VOLUME XXV NEW SERIES, NO. 171

JOURNAL
OF THE
AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER, 1930

CONTENTS

MEASURING JOINT CAUSATION. By Andrew T. Court 245

A WESTERNER'S EFFORT TO ESTIMATE THE POPULATION OF CHINA, AND ITS INCREASE SINCE 1650. By Walter F. Willcox 255

SOME APPLICATIONS OF STATISTICAL METHODS TO AGRONOMIC EXPERIMENTS. By Frederick D. Richey 269

SOURCES OF INCOMPARABILITY AND ERROR IN EMPLOYMENT-UNEMPLOYMENT SURVEYS. By Margaret H. Hogg 284

PRESENT CONDITION AND FUTURE PROGRESS OF STATISTICS. By Corrado Gini 295

NOTES:

Illustration of the Application of a Coefficient Measuring Correlation. By J. Arthur Harris and Borghild Gunstad 305

Wholesale Commodity Price Indexes. By Irving Fisher 308

Nomograms for Correcting Simple and Multiple Correlation Coefficients. By Harold D. Griffin 316

On the Theory of Contingency: I and III by Karl Pearson; II by J. Arthur Harris, Alan E. Treloar and Marian Wilder 320

The Program for the 1930 Annual Meeting. By M. C. Rorty 328

A Method of Appraising Short-Term Forecasts (333); Awards of the Social Science Research Council (334); Better Data on Merchandise Distribution (336); Cyclical Variations in Employment (339); Progress of the Census (344); Miscellaneous Notes (347); Members Added (360).

REVIEWS 363

PUBLISHED QUARTERLY BY THE
AMERICAN STATISTICAL ASSOCIATION
PUBLICATION OFFICE: Rumford Press, Concord, N. H.
EDITORIAL OFFICE: Columbia University, New York City

Price $1.50 per copy $6.00 per annum

THE EMPLOYMENT CLEARING HOUSE

OF THE

AMERICAN STATISTICAL ASSOCIATION

will aid you with your employment problem. It makes no charge to employers for its services. Any member of the Association may register for a position. There is no fee.

The following positions were open in August, 1930:

Field	Pay									
	Under $1800		$1800–2400		$2500–3600		$3700–4900		$5000 and over	
	M	F	M	F	M	F	M	F	M	F
Business-Mfg. and Selling			1							
Economic and Industrial Research	1				1		1		1	
Governmental							1			
Marketing					1					
Social Research						1				
Vital Statistics						1				

Members were looking for openings as listed:

Field	Pay									
	Under $1800		$1800–2400		$2500–3600		$3700–4900		$5000 and over	
	M	F	M	F	M	F	M	F	M	F
Advertising				1		1				
Agricultural Economics					2	2	2			
Business Mfg. and Selling	4	1	4	3	6	2	2		2	
Drafting			1							
Economic and Industrial Research	2	2	1	3		3	1	1	5	1
Finance	3		2		8	1	3		3	1
Insurance	2		2		1				1	
Labor					1			1		1
Marketing					1		1		2	
Mathematics		2	1						1	
Social Research			1	3		1				
Statistical Secretarial				1	1					
Teaching			1	1	1		1		3	
Vital Statistics		1			1	1				

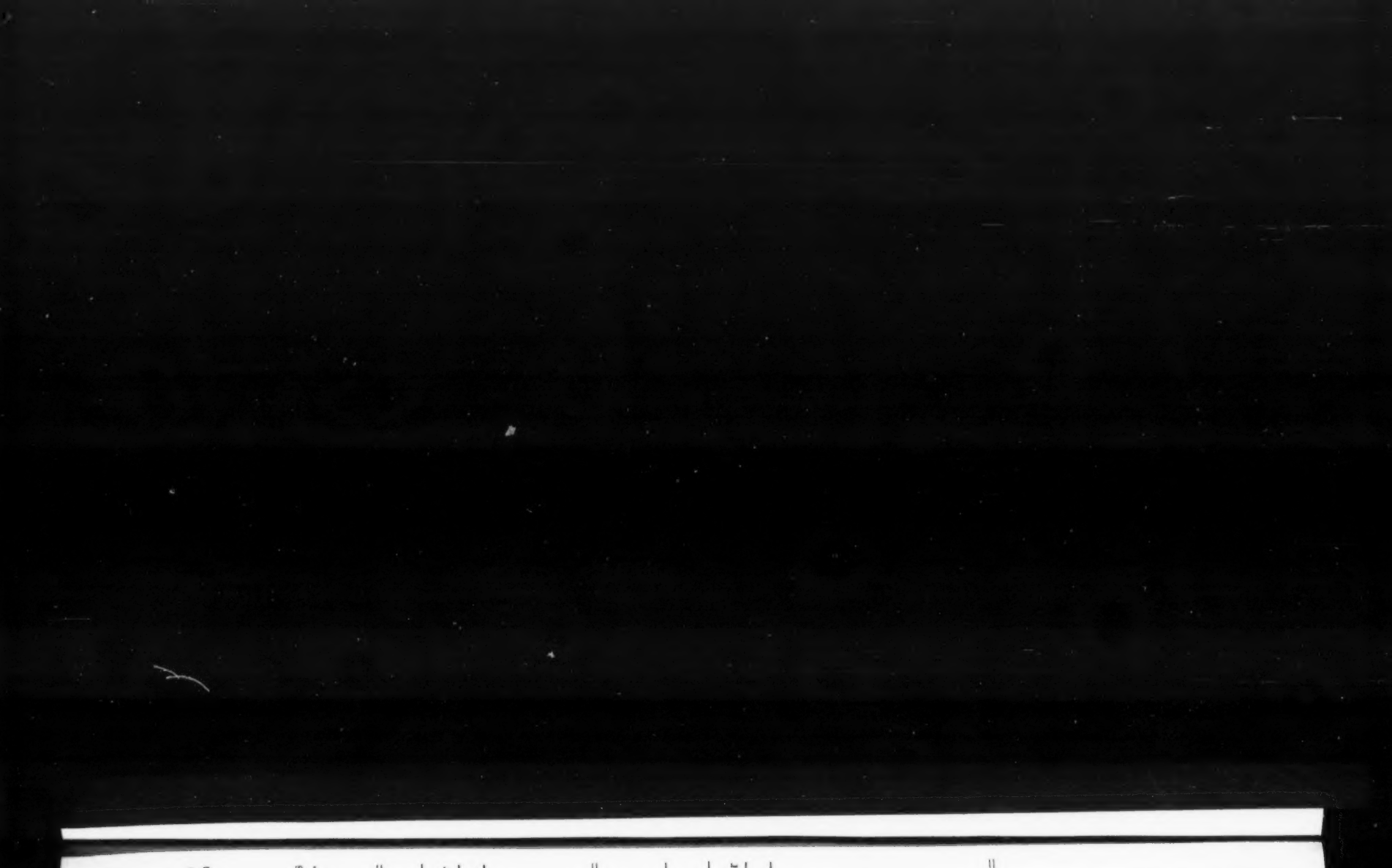
COMMUNICATE YOUR NEEDS TO

MRS. ROSE L. WOODBURY

Secretary's Assistant, American Statistical Association

Room 530, Commerce Bldg., 236 Wooster St., New York City

[NOT A PAID ADVERTISEMENT]



NEW SERIES, NO. 171 (VOL. XXV) SEPTEMBER, 1930

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Formerly the Quarterly Publication of the American Statistical Association

CONTENTS

MEASURING JOINT CAUSATION. By ANDREW T. COURT..... 245

A WESTERNER'S EFFORT TO ESTIMATE THE POPULATION OF CHINA, AND ITS INCREASE SINCE 1650. By WALTER F. WILLCOX 255

SOME APPLICATIONS OF STATISTICAL METHODS TO AGRONOMIC EXPERIMENTS. By FREDERICK D. RICHEY..... 269

SOURCES OF INCOMPARABILITY AND ERROR IN EMPLOYMENT-UNEMPLOYMENT SURVEYS. By MARGARET H. HOGG..... 284

PRESENT CONDITION AND FUTURE PROGRESS OF STATISTICS. By CORRADO GINI..... 295

NOTES:

ILLUSTRATION OF THE APPLICATION OF A COEFFICIENT MEASURING CORRELATION. By J. ARTHUR HARRIS and BORGHILD GUNSTAD..... 305

WHOLESALE COMMODITY PRICE INDEXES. By IRVING FISHER..... 308

NOMOGRAMS FOR CORRECTING SIMPLE AND MULTIPLE CORRELATION COEFFICIENTS. By HAROLD D. GRIFFIN..... 316

ON THE THEORY OF CONTINGENCY: I and III by KARL PEARSON; II by J. ARTHUR HARRIS, ALAN E. TRELOAR and MARIAN WILDER..... 320

THE PROGRAM FOR THE 1930 ANNUAL MEETING. By M. C. RORTY..... 328

A METHOD OF APPRAISING SHORT-TERM FORECASTS (333); AWARDS OF THE SOCIAL SCIENCE RESEARCH COUNCIL (334); BETTER DATA ON MERCHANDISE DISTRIBUTION (336); CYCLICAL VARIATIONS IN EMPLOYMENT (339); PROGRESS OF THE CENSUS (344); MISCELLANEOUS NOTES (347); MEMBERS ADDED (360).

REVIEWS: King: *The National Income and Its Purchasing Power*, by W. A. Paton (363); Hill: *Women in Gainful Occupations, 1870 to 1920*, by N. R. Deardorff (366); Brissenden: *Earnings of Factory Workers, 1899 to 1927*, by A. Achinstein (370); Hahn: *Die Statistische Analyse der Konjunkturschwingungen*, by S. Kuznets (372); Kohn: *Základy Teorie Statistické Metody*, by V. P. Timoshenko (373); Crum: *Corporate Earning Power*, by G. Soule (375); Lewinski: *Money Credit and Prices*, by M. A. Copeland (377); *Commission Internationale [illegible] la Revision Décenniale des Nomenclatures Internationales des Maladies*, by E. W. Kopf (379); King: *The Small Loan Situation in New Jersey*, by J. G. Rolph (381); Fisher: *Statistical Methods for Research Workers*, by H. Hotelling (381); *Joint Occupation Study*, by W. O. Menge (382); Cox: *An Appraisal of American Business Forecasts*, by E. C. Bratt (383).

Pubished Quarterly by the American Statistical Association

RUMFORD PRESS BUILDING, CONCORD, N. H.

Editorial Office, Columbia University, New York City

Entered at the post-office, Concord, N. H., as second-class mail matter

Acceptance for mailing at the special rate of postage provided for in Section No. 1103, Act of October 3, 1917, authorized April 29, 1922

Officers of the American Statistical Association

ORGANIZED NOVEMBER 27, 1839

PRESIDENT	Malcolm C. Rorty	
VICE-PRESIDENTS	Donald R. Belcher	Arthur R. Crathorne
	W. Randolph Burgess	Bradford B. Smith
	R. H. Coats	Edward L. Thorndike
SECRETARY-TREASURER	Willford I. King	
	Room 530, Commerce Building, New York University	
	236 Wooster Street, New York City	

CHAPTER AND DISTRICT SECRETARY

AUSTIN (TEXAS) CHAPTER
Carroll D. Simmons, *University of Texas, Austin, Texas*

BOSTON CHAPTER
Roswell F. Phelps, *Massachusetts Department of Labor and Industries, Boston, Mass.*

CHICAGO CHAPTER
H. B. Stair, *Illinois Bell Telephone Co., Chicago, Ill.*

CLEVELAND CHAPTER
D. C. Elliott, *Midland Bank, Cleveland, Ohio*

COLUMBUS (OHIO) CHAPTER
Harry S. Will, *66 Fallis Road, Columbus, Ohio*

LOS ANGELES CHAPTER
J. C. Clendenin, *639 South Spring Street, Los Angeles, Calif.*

SAN FRANCISCO CHAPTER
Merrill K. Bennett, *Food Research Institute, Stanford University, Calif.*

AREA REPRESENTED AND DISTRICT SECRETARY

DETROIT AND VICINITY
Lester K. Kirk, *Standard Accident Insurance Company, Detroit, Mich.*

PITTSBURGH AND VICINITY
George A. Doyle, *Bell Telephone Co., Pittsburgh, Pa.*

WASHINGTON AND VICINITY
Thomas B. Rhodes, *Federal Reserve Board, Washington, D. C.*

EDITOR	Frank Alexander Ross	
	405 Fayerweather Hall, Columbia University, New York City	
REVIEW EDITOR	Leo Wolman	
	30 Fifth Avenue, New York City	
ASSOCIATE EDITORS	Edmund E. Day	Wesley C. Mitchell
	William F. Ogburn	Walter F. Willcox

NEW SERIES, NO. 171 (VOL. XXV) SEPTEMBER, 1930

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Formerly the Quarterly Publication of the American Statistical Association

MEASURING JOINT CAUSATION

By Andrew T. Court

The multiple correlation equations in general use make an assumption which is often false in connection with the relationships they would measure. They assume that the net influence exerted by each of the independent variables can be expressed by a statement of relationship which ignores the simultaneous values of the other independents. They assume that for any change between given values of an independent variable there is a definite associated effect upon the dependent no matter what the values of the other independents are at the same observation.

This may not be the case with the organic relations which correlation methods are often called upon to define. The effect upon plant growth of a given change in temperature is not independent of the available supply of moisture. High temperatures that would stimulate in the presence of adequate water might mean death in a drought.

The equation chosen for use in multiple correlation analysis, be it mathematical or graphical, is almost always of the type

$$x_1 = f_1(f_2x_2 + f_3x_3 + \ldots\ldots f_nx_n).$$

This equation makes the explicit assumption that the dependent variable (x_1) can be explained satisfactorily in terms of some function of the algebraic sum of the functions of the independent variables. It assumes that the functions of the various participating causes combine only by addition to secure the observed effect. For many relationships this premise is not valid.

In spatial terms, surfaces of relationship reflecting joint causation are characterized by bending moments in directions not parallel to any axis, i.e., they are warped.

Dealing directly with this problem of measuring non-additive or joint causation and eliminating all of the limitations implied in and

necessary to any formal equation of relationship Ezekiel has devised a graphic and completely flexible procedure.[1] This method, which is essentially one of multiple smoothing and successive graphic approximation, has in turn been simplified by the suggestion of Waugh [2], who uses contours or isoropic lines for the orthogonal sections of Ezekiel. The philosophy and procedure of these two methods are essentially similar.

This paper explains a type of formal, empirical equation which has some of the flexibility of the freehand surfaces of Ezekiel and Waugh in measuring relations which show joint causation.

The empirical equation proposed here for use in measuring joint relationships is linear in its parameters and contains the following functions of the independent variables each of which is to be weighted in the presence of the other by the orthodox least squares procedure:

1. One or more compound elements constructed by combining two or more of the original independents into a single non-additive function which displays a marked degree of joint effect.
2. The independents in their original forms and also in individual functional modifications suitable for combination with the compound elements selected.

One such equation is

$$x_1 = ax_2 + bx_3 + c^{x_2} + d^{x_3} + ex_2^{x_3} + k.$$

The $ex_2^{x_3}$ factor provides the joint action while the other functions of the individual independents assist in orienting and modifying the total surface to fit the data.

Some experiments have been made with problems involving two independents. One of the simplest and most flexible equations of this type was found to be

$$x_1 = ax_2 + bx_3 + cx_2^2 + dx_3^2 + ex_2x_3 + f(x_2x_3)^2 + k.$$

The mathematical measurement of joint causation in this equation is accomplished by the inclusion of the compound elements (x_2x_3), and $(x_2x_3)^2$. The first of these (x_2x_3) is of peculiar value, for when used in combination with the linear and squared values of the individual independents it is capable of setting up parabolic bending moments at any or all directions not parallel to the axes. This possibility of bending at many angles provides a degree of flexibility and warp-

[1] M. J. B. Ezekiel, "The Determination of 'Curvilinear' Regression Surfaces in the Presence of Other Variables," this JOURNAL, Vol. XXI, September, 1926, p. 310.

[2] F. V. Waugh, "The Use of Isoropic Lines in Determining Regression Surfaces," this JOURNAL, Vol. XXIV, June, 1929, p. 144.

ability which enables the surface to define relationships which show joint causation.

Chart IV shows a surface which consists entirely of parabolic bending at an angle to the axes. This warping surface is one case of the above equation. This may be demonstrated briefly, for, if the following values described point O where the locus of this surface's minima intersects any line parallel to the Y axis, x_0, y_0, z_0 and tan O is the tangent of the angle between these two intersecting lines; and if x_1, y_1, and z_1 describe any point on the surface, besides a minimum, then the equation of this cylinder, whose section is a second degree parabola, is

$$z = z_0 + \frac{z_1 - z_0}{y_1 + \frac{x_1}{\tan O} - y_0 - \frac{x_0}{\tan O}} \left(y + \frac{x_0}{\tan O} - y_0 - \frac{x_0}{\tan O} \right)^2$$

which upon expansion is nothing more than one case of the general type

$$z = ax + bx + cx^2 + dy^2 + exy + k$$

which this method proposes.

The five series derived from two independent variables x, y, x^2, y^2, xy when weighted simultaneously according to the least squares criterion are capable of approximating not only any surface which has continuous bending moments parallel to the axes but also any combination of bending of any degree at any angle in between the axes so long as the general shape of the warping moment can be approximated throughout its observed range by one or a combination of several second degree parabolic cylinders with directrices perpendicular to a line parallel to the axis of the dependent variable. This provides a wide degree of freedom and at the same time keeps the number of variables reasonably small and the surface determined simple and smooth.

The warping surface built up from the weighted sum of the linear and squared independents plus the compound element consisting of the product of the two independents is capable of fitting itself at any angle or position without respect to the location of the origins of its constituent elements. It is, however, desirable from a mechanical point of view to orient the compound element itself in order to get full benefit of its adaptability for measuring joint relationships without undue refinement of calculation.

The surface described by the compound element alone approaches a plane closely at large values of x and y. This tendency is undesirable for it means that the non-additive effect becomes less and less important as a part of the surface and that the element becomes largely ex-

plicable in terms of $(fx+fy)$. There is an increasing degree of correlation between the original independents and the compound factor and, therefore, coefficients of net regression tend to become excessively large and unstable. It also means that in order to get full benefit of the joint effect inherent in the surface it is necessary to group very finely before correlation. This is undesirable.

To avoid this approach to a plane surface, it is well to insert arbitrary constants into the dependents, e.g.

$$(x-k_1),\ (y-k_2),$$

shifting them so that the observations in the study fall near the new origin before computing the compound element.

Beyond this shifting to use the most fruitful portion of the compound element [1] in the correlation, it is not necessary further to orient the factors. The combination and coöperation of the curved values of the original independents will twist the warping action of the compound element into whatever place the data dictate without further assistance.

To give some indication of the procedure and the possibilities of the method outlined above in the simplest terms, a problem was chosen in which the nature of the relationship to be measured had already been thoroughly investigated and well defined by non-correlation methods. The summary table of the effect of age and weight upon the mortality ratios of insured males which appeared in *Medico-Actuarial Investigation 1913*, Volume II, page 24, was selected. The surface defined by these "observations" (Chart I) is ideal for the purpose in hand because:

1. The surface displays a striking degree of joint causation. The variation of the dependent is not satisfactorily explicable in terms of the sum of any set of functions of the independent variables taken separately.

2. The "observations" are distributed uniformly over the surface, reducing the volume of computations to a minimum and still yielding good definition of the relationship to the very edges.

3. It has already been smoothed and refined with the interpolation formulae of actuarial science and presumably could be fitted almost perfectly with some empirical formula of reasonable complication, i.e., the purely erratic and meaningless deviations of the original observations have already been eliminated and the problem remaining for this analysis is simply that of testing the flexibility of the new technique described above against a ready-defined surface.

[1] This is accomplished automatically in grouping the observations to simplify computation of cross products. The series of cross products between the independent variables is the compound element (xy) sought, and needs only to be grouped before including it as an independent variable in the correlation.

CHART I

MORTALITY—BUILD—AGE
ACTUAL RELATION

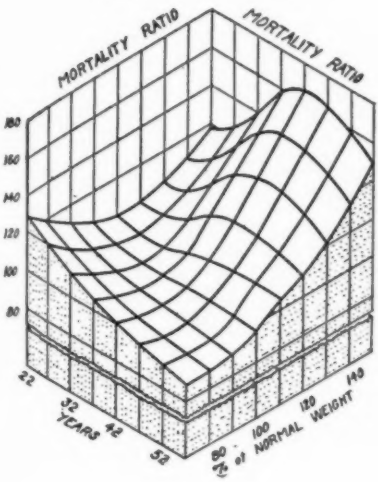


CHART II

M—B—A RELATION
ORTHODOX CORRELATION ESTIMATE

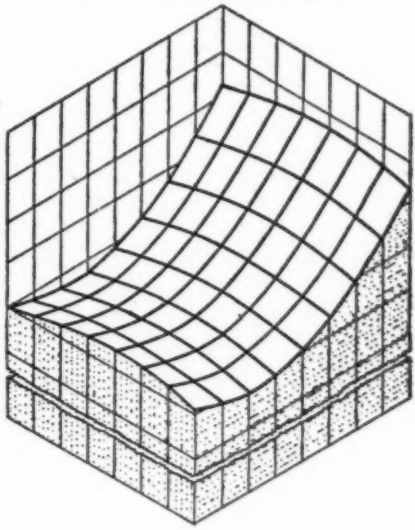


CHART III

M—B—A—RELATION
COMPOUND ELEMENT ESTIMATE

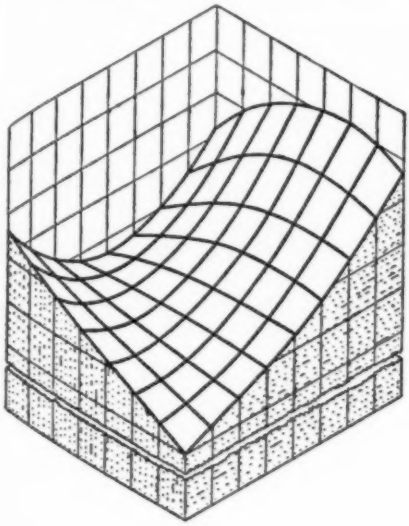
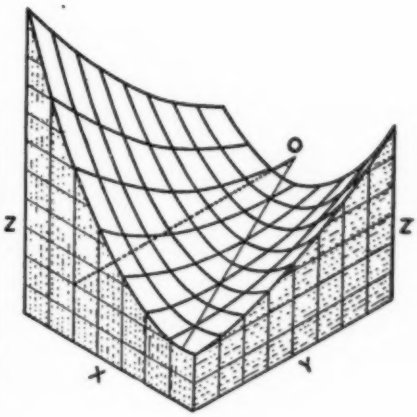


CHART IV

WARPING SURFACE



This build versus mortality table was rearranged, interpolated and grouped to divide ages into eight equally spaced groups, weight deviations into nine, and mortality ratios into twenty-three groups. See Tables I and II.

TABLE I

MEDICO-ACTUARIAL MORTALITY INVESTIGATION, 1913, MALES

Graduated percentages—actual deaths of expected by M.A. table according to percentage variation from average weight

Ages at entry	Percentage variation from average weight																
	−30	−25	−20	−15	−10	−5	Ave.	+5	+10	+15	+20	+25	+30	+35	+40	+45	+50
20–24	141	134	127	120	113	107	104	100	99	100	101	103	106	109	113	119	126
25–29	128	122	117	112	106	102	100	98	97	100	104	107	111	115	121	128	135
30–34	117	113	109	105	101	98	97	96	97	104	111	117	124	131	139	148	157
35–39	108	105	102	99	97	95	95	96	100	109	121	133	143	152	162	171	181
40–44	101	99	97	95	93	92	94	97	105	113	126	140	150	160	170	182	193
45–49	96	94	92	90	89	89	93	99	109	116	126	139	149	159	169	180	192
50–53	90	89	88	88	87	89	94	101	111	115	123	134	144	153	163	174	184
54–56	86	86	86	86	87	89	95	102	110	114	118	125	135	145	155	164	174
57–59	86	86	86	86	87	89	95	102	109	114	116	119	128	137	146	155	164
60–62	86	86	86	86	87	89	95	100	108	112	113	115	123	132	141	150	159

TABLE II

GROUPED MORTALITY DATA

x_2 \ x_3		1	2	3	4	5	6	7	8	9
1	x_1	12	9	7	5	4	4	5	7	9
2		10	7	5	4	3	5	6	8	11
3		7	6	4	3	3	6	9	12	15
4		6	4	3	3	4	8	13	16	20
5		4	3	3	3	5	9	14	18	23
6		3	2	2	3	6	9	14	18	22
7		2	2	1	3	6	8	12	16	20
8		1	1	1	3	6	7	10	14	18

$$x_1 = \frac{\text{Mortality Ratio} - 80}{5}$$

$$x_2 = \frac{\text{Age} - 17}{5}$$

$$x_3 = \frac{\text{Weight Deviation} + 40}{10}$$

The equation to be fitted to the data in question under the proposed method was [1]

$$x_1 = ax_2 + bx_3 + cx_2^2 + dx_3^2 + ex_2x_3 + f(x_2x_3)^2 + k.$$

The usual least squares procedure was employed for obtaining values for the parameters, a, b, c, d, e, f, and k, so that the sum of the squares of the deviations of the actual values of the dependent variable from the total estimate would be a minimum. The actual sequence of computation was, first, cross products between each of the series, second,

[1] No special orientation was necessary because the compound elements were computed from grouped data.

product moment coefficients [1] between each series, third, the Doolittle solution of the normal equations built up from the product moment coefficients and finally the estimates calculated from the net regression coefficients.

The actual estimates prepared from this last equation are shown in Chart III. This surface, it can be seen, reproduces most of the salient features of the original data and gives a fair picture of the variations in the effect of weight upon mortality at each of the different age groups.

This surface is still far from a perfect fit, but it does yield a significant and generally valid representation of the relationship under investigation.

The equation $x_1 = a_1x_2 + bx_3 + cx_2^2 + dx_3^2 + k$ was fitted to the coded observations according to the criterion of least squares and gave a correlation of .85. The surface defined by this orthodox equation is shown in Chart II. This surface fails entirely to indicate the significant changes in the relationship between weight deviation and mortality ratios which appear with the advancing years of life. In fact, the surface shows no appreciable variation between the relationships at different ages, and a simple two dimensional scatter of the mortality ratio against weight deviations would be substantially as informative. The correlation using the simple equation,

$$x_1 = ax_3 + bx_3^2 + k$$

is .84, practically as good as the more complicated multiple fit obtained above. Thus, the orthodox multiple correlation procedure would indicate that age is of no importance for the build-mortality relationship, but direct inspection of the original data gives the lie to that conclusion.

Any of these correlations could be further improved, of course, by applying the graphic regression procedure devised by Ezekiel [2] to the relation not only of the original independents with the dependent variable but also to that of the compound element with the dependent.

The table below compares in statistical terms the different correlation methods used in measuring the mortality relationship.

	R *Observed*	$\bar{R}$ *Corrected*[3]
Compound element analysis	.976	.973
Orthodox multiple correlation	.854	.842
Correlation omitting age	.844	.836

[1] The sum of the cross products between two series minus the product of the mean of one times the sum of the other.

[2] M. J. B. Ezekiel, "A Method for Handling Curvilinear Correlation for Any Number of Variables," this JOURNAL, Vol. XIX, December, 1924, p. 431.

[3] Corrected for the number of parameters relative to the number of observations. M. J. B. Ezekiel, "The Application of the Theory of Error to Multiple and Curvilinear Correlations," this JOURNAL, Vol. XXIV, March, 1929, p. 100.

In the course of a series of experiments which sought to test further the flexibility of the compound element method of measuring joint relationships, the data on potato yields in Maine used by F. V. Waugh in his recent paper [1] were analyzed by the proposed method. He has defined the surface of relationship between early and late rain fall and potato yield by freehand smoothing using isoropic projections. Charts V and VI represent, respectively, the freehand surface of Waugh, which may be considered as the ultimate shape any smooth surface would assume in fitting these data, and the mathematical surface

CHART V

MAINE POTATO YIELDS
FREEHAND ESTIMATE

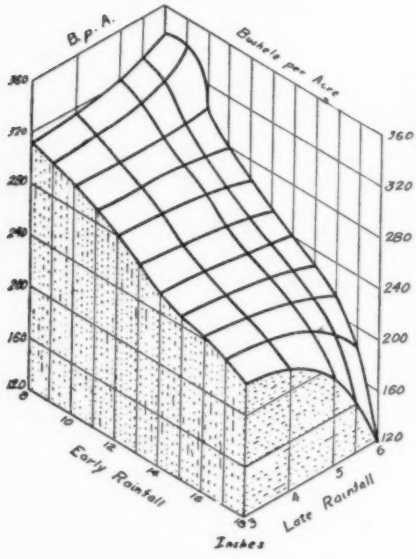
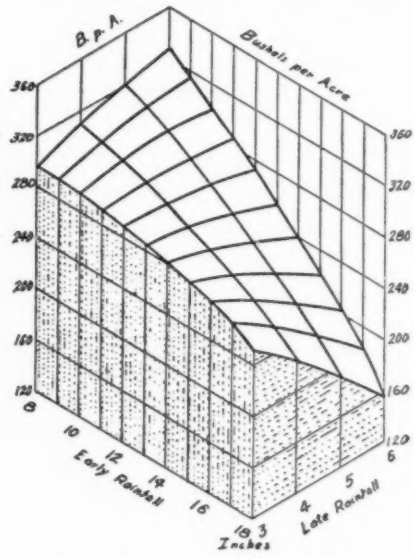


CHART VI

MAINE POTATO YIELDS
COMPOUND ELEMENT ESTIMATE



defined by the compound element equation when fitted to the original data.

It is significant that the mathematical equation although it agrees in general shape and slope with the freehand relationship does not conform to the purely local characteristics of the data. Other trials of this method have established the generality of this limitation. The compound element technique will measure a wide variety of rather complicated relationships in general outline. It cannot adapt itself to local peculiarities.

The potato yield problem was handled in a manner entirely analogous to the mortality work, except for the omission of the square of the prod-

[1] This JOURNAL, Vol. XXIV, June, 1929, p. 144.

uct of the independent variables. Since the number of available observations was so small it seemed desirable to limit the degrees of freedom in the equation as much as possible. The estimating equation when

x_1 = Maine Potato Yields
x_2 = Rainfall to August 1 (July doubled)
x_3 = Rainfall from August 1 to September 15

was (for 1928, after eliminating trend)

$$x_1 = 15.84x_2 + 59.07x_3 - .2758x_2^2 - 1.500x_3^2 - 4.2424x_2x_3 + 125.5$$
$$R = .94.$$

The possible uses and advantages of the compound element analysis may be listed:

1. The analysis can be made by routine arithmetic procedure with but little skilled direction. It provides a preliminary objective estimate of the general character of the relationship under investigation.

2. The method can cope with surfaces of more than three dimensions, directly by including in addition to $f(x_2x_3)$ such compound elements as $f(x_2x_4)$, $f(x_3x_4)$, and $f(x_2x_3x_4)$. Even where the relation is incapable of any spatial representation it is possible to proceed under routine methods. If it seems desirable later on to improve the fit of the original straight, curved and compound modifications of the independent variables, the actual work, no matter the number of interacting variables, reduces itself to a series of two dimensional relationship between the elements and the residuals after the least squares estimate.

3. The fact that every observation is included in the computation or in the graphical determination of the weight of each of the compound elements rather than splitting the data up into small independent groups falling in different sections of the surface makes it probable that the method often will give quicker and more accurate preliminary results than does the grouping multiple-way smoothing procedure, even in its isoropic modification. The accuracy of the ultimate surface depends upon the number of observations, the character of the relation, and the skill displayed in adjusting and refining the later estimates through freehand curved regressions. That accuracy will be pretty largely independent of the method used in the original analysis.

4. Since the number of independent variables used in the estimate is known it is possible to calculate the standard error of the coefficient of correlation and in general to estimate objectively the probable reliability of the results.

Among the limitations and disadvantages of this new compound element analysis are the facts that:

1. The analysis soon leads to a very large volume of computations as the number of variables increases. This is true however of any type of mathematical analysis and would necessarily apply with redoubled force to any generalized and flexible procedure such as the one proposed.

2. Some classes of relationship, e.g., periodic surfaces, would not yield to measurement with any reasonable number of compound elements of the proposed type. These relations are usually characteristic of mechanical rather than social or biological phenomena.

3. The approach through compound elements throws the results into a form not immediately interpretable in terms of actual surface shapes.

A WESTERNER'S EFFORT TO ESTIMATE THE POPULATION OF CHINA, AND ITS INCREASE SINCE 1650[1]

By Walter F. Willcox, *Cornell University*

The population of China must be determined in the end upon evidence obtained by the government of that country and accepted after analysis by scholars both in and out of China. But the time for that is not yet and may be a long way off. Since the seventeenth century, and especially of recent years, the problem has been studied by three small groups of scholars working almost independently. The first group is made up of Westerners who have lived in China, learned the Chinese language and familiarized themselves with Chinese sources, but are untrained in statistical technique and therefore ill qualified to weigh statistical evidence. The second group is made up of Westerners who have had little or no direct contact with China and cannot use even those few Chinese sources to which they have access, but are trained in statistics and have been attracted by this puzzling and important demographic problem. The third group is composed of Chinese scholars, few of whom are familiar with the statistical method and most of whom have examined the problem because of its relation to other social, economic or political questions of interest to China.

The aim of this paper is to review the provisional results thus far reached by the first and second groups and the evidence offered in support of them, so that they may be more easily criticised and thus perhaps another step may be taken toward the truth, for, as Bacon said long ago, "Truth cometh more easily out of error than out of confusion."

The paper which was presented before the Cairo meeting of the International Statistical Institute nearly three years ago[2] started with the assumption that the Chinese return of 1910, was probably more accurate than any other modern attempt to determine the population of that country. This assumption rested on the testimony of Tenney[3] and of Rockhill,[4] who had studied that return perhaps more carefully than other western scholars, and on the absence of opposing testimony. Within the last three years several scholars have indicated that they

[1] Read at the Tokio session of the International Statistical Institute, September, 1930.

[2] Walter F. Willcox, "The Population of China in 1910," this Journal, March, 1928, p. 18. Also in *Bulletin of the International Statistical Institute*, Vol. XXIII, Pt. 2, 1928, pp. 347–361.

[3] See United States Daily Consular and Trade Reports, July 13, 1911, pp. 184–187.

[4] W. W. Rockhill, "The 1910 Census of the Population of China," in *T'oung-Pao*, Vol. XIII, 1912, pp. 117–125. Also in *Bulletin of the American Geographical Society*, Vol. XLIV, 1912, pp. 668–673.

believe the assumption is not supported by affirmative evidence and is inconsistent with other evidence which they have brought forward. The main aim of the earlier paper, it may be explained, was not to exalt the return of 1910 but to answer the question, What is the interpretation of the figures in that return? A title such as "The Population of China According to the Return of 1910" would have been more exact than the one chosen "The Population of China in 1910."

Before discussing the evidence for and against the belief that the return of 1910 affords the best basis for estimating the population of China, the newly discovered evidence about the meaning of that return should be examined. The return of individuals, complete in four provinces and incomplete in two others, was then given under four heads: "males," "females," "school children" and "able-bodied males," but with no totals. The questions arise: (1) Is the population of a province to be found by adding all four classes, as was done in the first report from the American Embassy at Peking to the United States Government,[1] a report which has become the standard interpretation of that return? (2) Is it to be found by adding the first three groups and an estimated number of children under six years of age, as was done a year later by Rockhill [2] on the assumption that the first two classes included only adults and that the "school children" included all children over six years of age? (3) Is it to be found by adding only the first two groups?

The conclusion reached three years ago was that the first two groups were supposed to include not merely adults, but all males and all females. Neither of the Chinese words translated males and females carries an implication that it is limited to adults. The ratio between the number of "able-bodied males" and the number of "males" indicates that all males, even children under six years of age were included in the first class. In the earlier article was a statement, "If the instructions to the provincial authorities could be discovered, they might determine this crucial point." Through the kindness of Mr. W. W. Yen and the courtesy of the chief of the Bureau of Statistics in Peking they have been discovered and my friend, K. P. King, has furnished me with a translation of them. In the instructions for the "enumeration of persons" is a sentence, "the children at seven years [i. e. 7 and over] and male persons at sixteen [i. e. 16 and over] should be separately presented at the end of the report." This seems to confirm the earlier interpretation that these two special classes although included in the first two columns were also reported separately. Perhaps the Chinese

[1] See footnote 3, *supra*.

[2] Rockhill, *op. cit.*

authorities desired especially to learn the number of children of school age and of men able to bear arms.

A more fundamental question is: Do the 1910 returns furnish the best basis for estimating the present population of China? Against the belief that they do, arguments have since been raised centering around two questions: (1) Was the second part of the return of 1910 vitiated by serious omissions, especially of very young children and of females of greater age? (2) Was the return, as a whole or at least the second part of it, more reliable than previous or subsequent estimates or reports?

There is some evidence that children under two years of age or thereabouts were omitted. A valued correspondent, who has had experience with local enumerations in China, has written to me from Nanking, "Babies are usually not enumerated by Chinese when giving the number of persons in the family. Within the meaning of the Chinese term for persons, babies do not qualify." To the same effect is a statement that "all the people, from the babe who has teeth up to the man are recorded in the census." [1] The fact that in the incomplete population return following upon the return of households only four-fifths as many females as males were reported, a difference in the proportion of the sexes much greater than that in any other country and about the same as in Montana or Wyoming, arouses suspicion of the figures. This last difficulty might be met by supposing that female infanticide is prevalent enough in China to produce this marked disproportion in the number of the sexes. A bit of newspaper evidence may be mentioned as pointing in that direction. In the city of Shanghai a Public Benevolent Cemetery Society was organized in 1928 with Chinese officers to provide among other things "for burial of indigent or unknown dead persons," and in 1929 it found and cared for the bodies of more than 28,500 dead infants. The implication of the newspaper statement before me is that most of these were "unwanted girl babies."[2] The population of Shanghai is estimated at approximately 1,500,000. If it be given a birth rate of 30 per thousand, and in view of the excess of males apparently characteristic of most cities in the far east it would not be wise to put the rate higher, the annual number of births would be about 45,000 or 22,000 females and 23,000 males. If many more than half of these dead infants were females, it would seem fair to infer that female infanticide is practised in Shanghai on a great scale. If the same is true also through large parts of China the accuracy of the return of 1910 by sex might be defended.

[1] Chen Huan-Chang, *Economic Principles of Confucius*, 1911, Vol. I, p. 298.

[2] *The New York Times*, March 8, 1930. Special Correspondence from Shanghai, dated January 30.

But if that return was inaccurate the probable cause is the failure to report many females. To estimate the population from the 1910 return on the assumption that all children under two years of age and many females were omitted, the true proportion of children under two years of age to the rest of the population and the true proportion of females to males in the population over two years of age in China in 1910 may be supposed to have been the same as they were in Japan in 1908. On these assumptions the true number of females over two years of age in those parts of China for which the return of persons was completed, was 44.7 million instead of 36.4 million and the omitted children under two years of age numbered 5.4 million. If these two changes are adopted, the population is increased from 82.0 million reported to 95.7 million or by 16.7 per cent. It seems probable that the proportion of females is less in China than in Japan, due to neglect of female children in the former country, and that not all children under two years of age were omitted in the return of 1910. If so, the true population in the area under consideration was more than 82.0 million and less than 95.7 million. With the use of these assumptions, the following results in thousands are reached for the four provinces:

Province	Number of households	Estimated population	Persons to a household
Chihli	5,208	30,000	5.8
Chekiang	3,888	15,070	3.9
Kiangsi	3,440	16,890	4.9
Kweichow	1,772	9,920	5.6
Total	14,308	71,880	5.0

If these results for the four provinces are accepted and extended to the old eighteen provinces of China Proper, they would indicate a population of 323,000,000 greater by 55,000,000, or one-fifth, than the unmodified results of the return of 1910 as interpreted three years ago.

The foregoing figures, however, show surprising differences between the provinces in the average size of a household. Why is a household in Chekiang only two-thirds as large as in Chihli? The range in these four provinces is greater than in the 48 states of the American union in 1920, and yet the difference between North Carolina and Nevada in size of household one would expect to find greater than that between Chekiang and Chihli. In view of its exceptional character and its emphatic rejection by persons who have lived long in that part of China and studied the population problem on the ground, the figure for Chekiang Province should perhaps be disregarded. If the changed figures for the other three provinces are accepted, the average size of

a household in them is 5.4 instead of 5.0. Three recent studies of the farm population in certain selected districts of China embracing more than 14,000 families and 75,000 persons have shown 5.2, 5.3 and 5.7 persons to a household, or in combination 5.3. The authors of these studies believe that, notwithstanding the care with which their field work was done, some very young children were omitted. If all children under two had been omitted and if the proportion of such children to their elders had been equal to the average in Japan in 1908, the size of a household would have been 5.6 instead of 5.3 persons. A comparison between the Japanese figures of 1908 and those in one small group of farm families for which the age distribution of the reported children under five is given, suggests that about two-fifths of them were missed in the farm study. The conflicting evidence indicates that the average size of a household in China lies between 4.3 and 5.4 persons and somewhat nearer the upper limit than the lower. It is best perhaps to use 5.0 as a multiplier.

If the return of individuals in 1910 was untrustworthy, is there reason to believe that the return of households was more accurate? The former was an independent inquiry following that of households by about a year and novel in character, while returns of households had been made many times and both local officials and the public had become accustomed to them. The fact that the return of individuals broke down and was left incomplete gives perhaps another reason for looking with suspicion upon such returns as did come in. It seems likely that the return of households, while far from trustworthy, was better than that of individuals. If that return is accepted and the average size of a household is assumed to be 5.0 persons, Table I results.

This table assigns China about 48 million more people than were indicated by the earlier interpretation of the return of 1910. The difference results from rejecting the average size of household shown on the face of its figures and accepting the interpretation that all children under two years of age, or an equivalent number of very young children, and many million older females were not reported. The evidence on which Rockhill relied in 1904[1] for an average of four persons to a household and which at first it seemed best to accept on his authority does not bear analysis, because some at least and possibly all of the returns he cites as indicating a small household seem to have been returns of adults or adult males, rather than, as he thought, of all persons.

It may help successors who wrestle with this problem, especially

[1] W. W. Rockhill, "An Inquiry into the Population of China," in *Report of the Smithsonian Institution for 1904*, Washington, 1905, pp. 659–676.

TABLE I

HOUSEHOLDS AND POPULATION OF CHINA PROPER AND ALL CHINA (IN THOUSANDS) BY THE RETURN OF 1910 AS REVISED

Province	Households	Population on the basis of 5.0 persons to a household
North China	*15,095*	*75,475*
Chihli	5,208	26,040
Shantung	5,380	26,900
Shansi	1,993	9,965
Shensi	1,605	8,025
Kansu	909	4,545
Central China	*25,878*	*129,390*
Kiangsu	3,148	15,740
Anhui	3,148	15,705
Honan	4,662	23,310
Hupeh	4,533	22,665
Szechwan	10,394	51,970
South China	*23,566*	*117,830*
Chekiang	3,888	19,440
Fukien	2,379	11,895
Kiangsi	3,440	17,200
Hunan	4,288	21,440
Kweichow	1,772	8,860
Yunnan	1,572	7,860
Kwangsi	1,175	5,875
Kwangtung	5,052	25,260
China Proper	*64,539*	*322,695*
Outlying Districts:		
Sinkiang	*467*	*2,335*
Manchuria	*2,580*	*12,900*
Fengtien	1,600	8,000
Kirin	739	3,695
Heilungkiang	241	1,205
Mongolia	....	*1,800*
Tibet	....	*2,000*
Total Outlying Districts	....	19,035
China Proper	....	322,695
All China	....	341,730

those in the Far East who may not have ready access to Western sources, if a summary of the development of Occidental opinion and argument about the population of China is included in this report.

The earliest statement about that population in any European source, goes back to the sixteenth century and was written by a man sometimes called the father of statistics. It runs thus:[1]

> There are said to be in China one hundred and fifty Cities, two hundred thirty-five great Townes, one thousand one hundred fifty foare Castles, and foare thousand two hundred Boroughs without walls . . . of Villages and Hamlets (some of them containing a thousand households) the number is infinite: for the Country is so covered with habitation that all China seemeth but as one Towne. . . . Of the number of the Inhabitants, no certainty can be produced, but according to manuscript relations and the report of travellers, it is said that the Kingdome containeth threescore and ten millions of living soules.

[1] Giovanni Botero, *Relazioni Universali*, Rome, 1593. The quotation is from the English translation, London, 1630.

In the middle of the following century, more exact information about the population of China reached Europe through Father Martin Martini, who had spent years in that country and after his return published his *Novus Atlas Sinensis*, the geographical and statistical facts in which he drew "from their own antient records ever since the time of Noah." The book contains a careful and detailed description and map of each of the fifteen provinces of China and gives the number of households and of men in each. In the introductory chapter the following puzzling passage occurs:[1]

> The most authoritative books about China if we credit them noting with care and accuracy the number of men [hominum] who were in each province, city or place (omitting those of the royal line, magistrates, eunuchs, clergy or priests of their superstitions, women and children) yield a total of 58,914,284. Hence it is not surprising that China is said to have 200 million men [hominum].

The problem is to understand how Martini reached, if he did reach rather than merely report, the figure of 200 million. Could he have told some friend versed in Chinese lore that Europe was thought to contain about 100 million people and been gravely assured that China contained fully 200 million? The households in each province, by his figures, total 10,128,787 and the men "homines" 58,916,783. If the last figure is interpreted as the total population, there would be 5.8 persons to a household. On the other hand, if the figures of 200 million persons and about 10 million households are accepted, that would mean about 20 persons to a household. One other passage in the *Atlas* besides the second line above includes homines meaning adult males, but in more than a dozen cases it means apparently population. Martini seems to have been regarded until into the nineteenth century as the most authoritative western writer upon Chinese questions, but his estimate of the population was criticised acutely in 1740 by Struyck, who wrote that the 58,916,783 probably included all the men, women and children in China except such as were not subject to the census and rejected the figure of 200 million because it would mean that the household included about 20 persons and about 5 males over twenty years of age.[2] Süssmilch, a contemporary of Struyck, discussed the population of China at greater length but with less penetration and was disposed to accept the population of 200 million.[3]

Volney in 1803 set the population of China at about 120 million and as authority mentioned an [unidentified] enumeration published a year before by the English and showing a population of only 55 million in

[1] Martin Martini, *Novus Atlas Sinensis*, Amsterdam, 1655.

Nicolaas Struyck, *Inleiding tot de Algemeene Geographie*, Amsterdam, 1740, pp. 323 ff.

J. P. Süssmilch, *Die Göttliche Ordnung*, 1762, Vol. II, pp. 218–220.

the country districts of China.[1] He supposed that the Chinese cities had an equal population (?) and thus reached a figure of 110 million which he then raised without explanation to 120 million.

Malte-Brun wrote in 1810:[2]

> Among the different accounts respecting China how shall we hit upon the true one? Has that country 27,000,000 inhabitants according to Sonnerat? or 55,000,000 according to the extract from the *Official Gazette* of Pekin? or 70,000,000 according to the Russians? or 100,000,000 as de Guignes believed? or 19,662,000 as Mr. Buesching informs us? or 200,000,000 agreeable to the statement of the missionaries? or, finally 333,000,000 as a Chinese mandarin, perfectly veracious no doubt, assured Lord Macartney?

Worcester in 1817 took from Georg Hassel, whose work I have not found, a table showing the population of the Asiatic countries and giving China 150 million.[3]

Wells Williams in 1848 examined the subject more carefully than any previous western writer. He had had the advantage of long residence in China and acquaintance with Chinese censuses which he says "with all their discrepancies and inaccuracies are the sources of information upon which the most dependence can be placed." He relied upon the return of 1812 which he called "the most definite and satisfactory of all the censuses in Chinese works." It showed a population of 362 million.

Dieterici, in 1859 made the first modern and significant effort to estimate the earth's population and adopted Williams' figure for China with a slight modification which raised it to 367 million.[4] Sacharoff, perhaps the most careful and thorough western student of Chinese population, reproduced in 1858 the Chinese return of 1842 showing 415 million people.[5] He did not state what figure he thought most probable, but evidently believed the Chinese returns excessive. About the estimate of 333 million reported in 1872 by Commissioner Chau to Macartney, the first Ambassador from Great Britain to China, Williams wrote that it had received much more respect than it deserved and that it was "the off-hand remembrance of an intelligent native."[6]

[1] C. F. Volney, *Tableau du Climat et du Sol des Etats-Unis d'Amerique*, 1803, Vol. II, pp. 474–477, footnote.

[2] Malte Brun, *Précis de la Géographie Universelle*, Paris, 1810. Translated and published, 1824, Vol. I, p. 549.

[3] J. E. Worcester, *Gazetteer*, 1887, Vol. II, end.

[4] K. F. W. Dieterici, "Die Bevölkerung der Erde," in *Petermann's Mittheilungen*, 1859, pp. 1–19.

[5] T. Sacharoff, "Historische Uebersicht der Bevölkerungsverhältnisse Chinas," in *Arbeiten der kaiserl. russischen Gesandtschaft zu Peking über China*, Berlin, 1858, Vol. II, pp. 129–195. Quotations are from the translation by W. Lobscheid published at Hongkong in 1864.

[6] S. Wells Williams, *The Middle Kingdom*, 1848, Vol. I, pp. 206–234.

The first volume of the *Geographisches Jahrbuch*, appearing in 1866, gave a return of 1852 showing 537 million but preferred an estimate of 450 million reached by carrying forward the increase between 1812 and 1842.[1] The third volume apparently reduced the estimate to 420 million because of the losses from civil war, famine, and pestilence.[2] In 1872 the *Jahrbuch* was merged in *Die Bevölkerung der Erde*. After considering evidence offered by various travelers, that publication raised the earlier figure to 447 million.[3] In 1874 it reported that persons acquainted with China were almost agreed upon 400 million as the best estimate of the population so it reverted to the return of 1842 with changes in four provinces notably Chekiang due to later evidence, and reached a total of 405 million.[4] In 1882, apparently as a result of a growing volume of opinion that the population of China was less than current estimates, it decided that the return of 1812 was better than that of 1842 and adopted the former after changing the figures for a few provinces because of new evidence, thus reaching a total of 350 million.

In 1887 Levasseur and Bodio, estimated the population of China at 383 million.[5] This they did by abandoning the return of 1842 and using smaller figures, some for 1882 and others for 1879.

In 1891 *Die Bevölkerung der Erde* examined the then recently published comments of Behm, Temple, Richthofen, Young, Hopper, Hippesley, Ratzel and Mayet and proceeded to a study of the returns for 1812, 1842, 1879–82 and 1885, province by province, computing the density of population which would result from the reported figures and weighing its probability.[6] This careful analysis led the editor to an estimate of 350 million, much below the 413 million reported for 1842 and slightly below the 360 million reported for 1812.

Lastly in 1901, the same publication under a new editor, Supan, renewed the attack upon the same problem by comparing the series of Chinese figures with one another in the light of the external evidence at hand.[7] He accepted Sacharoff's statement of 1858 borne out by his and Parker's evidence that in 1775 the figures were deliberately padded by local Chinese officials in order to please the Emperor who had indicated a desire for a larger total. The increase in one year as given by Parker, 43.5 million or 20 per cent, the editor believed was purely fictitious and had been retained in the record for more than a century.

[1] *Geographisches Jahrbuch*, edited by E. Behm, Gotha, 1866, Vol. I, p. 63.
[2] *Idem*, Vol. III, 1870, p. 57.
[3] *Petermann's Geographische Mittheilungen*, Ergänzungsheft No. 33. "Die Bevölkerung der Erde," pp. V and 38–39.
[4] *Idem*, pp. V and 40–41.
[5] *International Statistical Institute, Bulletin*, Vol. II, Pt. II, 1887, p. 191.
[6] "Die Bevölkerung der Erde," VIII, 1891, pp. 98–105.
[7] *Idem*, XI, 1901, pp. 44–50.

Sacharoff explained the increase thus:[1]

> This consideration . . . induced the government to show the people of China unheard-of favor. In A. D. 1712 an edict was issued which decreed that the number of taxpaying people in A. D. 1711 should forever remain the standard for those subject to crown service; the rest of the population . . . being exempted from payment of taxes. . . . Magistrates were ordered . . . to take the census of the population . . . every year. . . . The authorities . . . seeing that a strict supervision of the census was no longer a source of profit; and considering the same of no use to the government, entirely neglected this duty or left its execution to the caprices of their inferiors. . . . The time of registration approaches and the local authorities, afraid of their responsibility and fearing a decrease of the population, copy the tablets without inquiring after the presence of the persons put down in them. . . . Upon the arrival of the time of general verification . . . the local authorities generally commission their own clerks who are less anxious to ascertain the real number of inhabitants than to ease their own labour. Hence an incredible number of temporary residents such as travellers, mechanics, etc., find an opportunity of entering their names at two or three places . . . A. D. 1775 the Emperor himself, by comparing the lists sent to him by the Governor-General with those of former years detected great negligence and lack of attention in the compilation. . . . He issued an edict commanding that a census be taken in the most careful and reliable manner. . . . The observations of the Emperor respecting the small increase induced the authorities to take greater care for the future. They invented another mode of compiling the lists by arbitrarily augmenting the figures . . . thus showing a steady increase of the population.

By a complicated, ingenious and plausible method Supan endeavored to eliminate this century old padding and in that way reduced the previous estimate of 350 to 320 million which is apparently the best interpretation of Chinese returns made by a Western scholar.

Let us now ask what seem to be the best guesses, they are hardly more, at the population of China Proper or the old eighteen provinces in 1650, 1750, 1850 and today and so what has been the probable rate of increase?

What was the probable population of China Proper in 1650? Rockhill inferred from certain Chinese statistics, "that the total population of China Proper in 1651 . . . was about 55 million . . . just about the number we should have assumed it to be had we to deduce it from the data supplied by history alone."[2] The discussion in Williams leaves the impression that he would estimate China's population in the middle of the seventeenth century at about 60 million. Sacharoff estimated the population of China in 1661 at 21 million. Gowen wrote, "According to the accounts given by the Jesuit fathers, China at this

[1] Sacharoff, *op. cit.*

[2] W. W. Rockhill, "An Inquiry into the Population of China," in *Report of the Smithsonian Institution for 1904*, Washington, 1905, pp. 659–676.

time (1644 A. D.) contained a population of 11.5 million families. At the end of the reign (1661 A. D.) another estimate was made of nearly 15 million families or 89 million individuals." Parker compiled from the annual volumes of the official *Tung-hwa-Luh* the numbers of taxed families for each year in the last half of the seventeenth century.[1] Disregarding the figure for the first year because it is much smaller than any other, the average number of such families, 1652–4, was 14,153,000, pointing to a population of about 77 million. Chang-Heng Chen, relying apparently on the same Chinese sources, has stated that during the 60 years of Kang Hsi's reign (1662–1722) the census population "limited to males between 16 and 60 for taxation purposes" stood uniformly at about 27,350,000.[2] If this figure is correct and if at that time in China, as today in Japan, the males 16–59 years of age were 27.6 per cent of the population, China had a population of about 100,000,000. But if the improbable idea that the population was stationary during those 60 years of prosperity is abandoned and Chang-Heng Chen's figure of 27,350,000 males between 16 and 60 be accepted only for the middle of the reign and the rate of increase of population between 1692 and 1749 be carried back to the year 1650, then the population at that date was about 66 million.

From these various sources the following estimates of the population of China about 1650 have been derived:

	Millions
Sacharoff	21
Rockhill	55
Martini	59
Williams	60
Chang-Heng Chen (adjusted)	66
Gowen	69
Parker	77
Chang-Heng Chen	100
Martini	200

Reasons have been given for rejecting Martini's larger figure. It seems likely that Sacharoff's figure was erroneous because he accepted a return of adult males as one of the whole population. If these two extreme estimates are discarded the average of the other seven is about 70 million, which may be adopted as the best approximation to the population of China in 1650.

What was the probable population of China in 1750? Parker took the figure of 180 million from an official Chinese document, and Chang-

[1] E. H. Parker, "A Note on Some Statistics regarding China," in *Journal of the Royal Statistical Society*, Vol. LXII, 1899, pp. 150–156.

[2] Chang-Heng Chen, "Changes in the Growth of China's Population in the Last 182 Years," in *Chinese Economic Journal*, January, 1927.

Heng Chen using apparently the same source reached substantially the same result. If the previous estimate of 70 million in 1650 is accepted, this increase of 157 per cent in a century, nearly one (0.95) per cent a year, or about two-thirds as rapid as the annual rate of increase in the United States between 1910 and 1920 (1.45 per cent), seems surprisingly but perhaps not incredibly high.

The count of households in China has probably been more accurate than the count of heads. Parker gave from the *Tung-hwa-Luh* the number of families for nearly every year between 1651 and 1734. Disregarding the very low figure in the first return the number in 1653 has been estimated as the average for 1652–4, and that for 1733 as the average for 1732–4. The increase for that four-fifths of a century was about 82 per cent or three-fourths of one per cent a year. If the increase of households in the other fifth of the century was at the same rate as in these four-fifths, and if the increase of population was at the same rate as that of households, then a population of 70 millions in 1650 would have become 148 million in 1750.

If the two centuries between 1657 and 1851 be divided at the year 1707, a notable difference appears between the rates of increase of the Chinese population in the two periods. The average annual rate of increase in the century and a half after 1707 (0.94 per cent) was more than five times as great as that (0.19 per cent) in the half century before 1707. It is difficult to credit these Chinese figures showing a fourfold increase in about a century and a half. To confirm them there is need of evidence, not only that the period was one of peace and prosperity, but also that the available food supply increased rapidly or the consumption of food per head and the average standard of living sharply decreased.

Sacharoff's statements already quoted derive some confirmation from a table in which he gives the population of each province of China at each of ten returns between 1749 and 1842. The table gives 171 cases of change between two successive returns, six of which are apparently decreases and 165 increases. But Supan has pointed out that four of these apparent decreases are due to mathematical or typographical errors, so that on the face of the figures there were 169 increases and only two decreases of population in the provinces of China between 1749 and 1842. An increase of population, so unbroken in time and spread over so extensive an area densely settled at the start, is almost if not quite unparalleled in the annals of mankind. Perhaps when the history of that period is better known (the Manchu annals are still for the most part in manuscript) and if it can be shown that Chinese agriculture was extended into new and fertile districts or introduced new and more

productive crops or better methods of tillage, the Chinese figures will come to be accepted. For the present it seems wiser to reserve judgment upon them.

If a provisional opinion about the modern growth of Chinese population may be offered for discussion, it would include the following points (1) reject all official figures for the period between 1711 and 1912, when they seem to have been used not for any administrative purpose but only to give pleasure to the Emperor and his gods, (2) interpret the return of 1910 as indicating a population of 342 million for all China and 323 million for China Proper, (3) treat that return as more trustworthy than later and larger official figures, (4) in view of the lack of reliable returns between 1850 and 1910 for all China and of the many natural and political calamities from which the country suffered during the later years of Manchu rule, assume that the population of China in 1850 was about the same as in 1910, viz., 342 million, (5) accept Parker's figures drawn from an official Chinese source, that the number of households in 1710, just before the decree of 1711, was 23,311,000 and his conclusion that the household then averaged about 6 persons, indicating a population of about 140 million. With the help of these assumptions, and of the earlier results, the population of China Proper at various dates may be set provisionally, and pending the production of further evidence, as follows:

Date	*Estimated population of China Proper (in millions)*
1650	70
1710	140
1850	342
1910	342
1930	342

If these rough estimates are accepted and it be assumed that the population increased between 1710 and 1850 at a uniform annual rate, the population of China in the middle of the eighteenth century was about 181 million. This is nearer the Chinese official figure of 180 million than the 148 million reached by continuing for 17 years the average rate of increase of households between 1653 and 1733.

These results are to be regarded not only as subject to a very large probable error but also as purely provisional. The outstanding fact is that no one knows what the population of China is within many millions. The aim of scholars should be to get nearer the truth than they now are. May I close with a few suggestions about lines of work which might help our successors to better results?

1. Most Western scholars are handicapped by inability to read Chinese. Yet when the population of China is determined it will be by the critical analysis and interpretation of Chinese sources old or new. A compilation, made and published in English or some other western language, giving all the information about the population of China to be found in Chinese works is much needed. This should be a set of figures not merely for all China but also for each province and if possible for each Hsien or county at the date of each return. Explanations should be given about the method followed. At present work is being repeated because later writers are not properly informed about the results of their predecessors. For example, E. H. Parker spent three years in going through the 70 volumes of the *Tung-hwa-Luh* so as to find "any casual remarks bearing upon population and revenue," and in copying out the numerical details about the annual returns on these subjects. But the table which was the final result of this toil was not included in any book of his and has been overlooked by his successors.

2. Chinese scholars also are handicapped, but in other ways. They know almost as little of what Western scholars have written as the latter do of what Chinese scholars have written. The modern statistical method is mainly a Western development. A survey of the Western discussions about the population of China, at which I have made a beginning, would indicate what kind and degree of evidence are needed to satisfy the guild of scholars.

3. A summary of the areas of the Chinese provinces and Hsien or counties is needed so that the density of population in small districts of China can be computed and compared with densities elsewhere. The evidence suggests that the richer agricultural districts of China, especially the rice-growing districts, are very well settled, but that in the less productive upland and inland regions, where unaided hand labor can wring little from the soil, the population is comparatively sparse. Densities both per square mile of total area and per square mile of tilled land are much needed.

4. The methods available for estimating the population of unenumerated countries described years ago by a committee of the International Statistical Institute, would probably be applicable and helpful.

SOME APPLICATIONS OF STATISTICAL METHODS TO AGRONOMIC EXPERIMENTS[1]

By Frederick D. Richey, *Bureau of Plant Industry, U. S. Department of Agriculture*

One of the most general uses of statistical methods in agronomy is the estimation of the experimental error, or error due to random sampling, in field plot experiments. Most agronomic experiments must be conducted in the field and are subject to the errors of random sampling among variable plants growing in a variable soil. If many different strains or varieties are to be compared, for example, it is desirable to measure the precision of the comparison. The limitations of land and labor preclude replicating enough to obtain a small mean experimental error merely by increasing the number of replications. In fact, the decrease in the standard error of the mean which theoretically accompanies an increasing number of replications very frequently does not occur. This is due to the inclusion of greater soil variation as the experimental area is enlarged to accommodate the greater number of replications. Moreover, much of the soil variation is not random, but comprises systematic change from one part of the field to another. Some means of obtaining a reasonably reliable standard error based on a small number of replications of any particular item, and of having such standard error as free as may be from the effects of systematic variation, accordingly is exceedingly important.

The data in Table I show the yields of McCulloch corn grown in a uniform-planting experiment, at Ames, Iowa, in 1925. They are the yields of 192 12-hill plots, all of the same kind of corn, arranged in four series of 48 rows each, just as they occurred in the field. The row numbers are shown in column 1 of the table and are common to the four parallel series, or replications. It has been assumed that the experiment was planted with 48 different strains of corn, numbers 1 to 48, and that these were distributed at random in each of the four series, with the single restriction that each of the 48 strains should occur once and only once in each group of 12 rows or plots across the 4 series. Columns 2 in the table show the assumed strain number for each plot. The yields of air-dry corn are shown in columns 3. The problem is to estimate the probable significance of the differences between the means of the various "strains."

[1] Presented in part before the Round Table on Biology, at the annual meeting of the American Statistical Association, Washington, D. C., December 30, 1929.

Forgetting for the moment the assumption that the experiment includes different strains, the standard deviation for the experiment, of course, may be estimated by squaring the deviations of the individual plots from the general mean, adding, and dividing by 191, one less than the total number of plots. Designating the value so obtained as the total standard deviation (σ_T), it is found to be 1.999.

TABLE I

YIELDS OF 192 12-HILL ROWS OF McCULLOCH CORN AT AMES, IOWA, IN 1925, ASSUMED TO BE THOSE OF 48 STRAINS DISTRIBUTED AT RANDOM WITH RESTRICTIONS

(Data from A. A. Bryan)

	Series I				Series II				Series III				Series IV			
Row	Strain	Yield	Dev. from strain mean p.c.	5-row mov. av. p.c.	Strain	Yield	Dev. from strain mean p.c.	5-row mov. av. p.c.	Strain	Yield	Dev. from strain mean p.c.	5-row mov. av. p.c.	Strain	Yield	Dev. from strain mean p.c.	5-row mov. av. p.c.
1	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
1	36	14.96	19.89	21.00	11	15.17	21.20	20.81	35	14.33	20.75	19.90	8	16.12	20.61	21.14
2	34	14.54	21.02	21.00	19	13.94	20.32	20.81	31	15.65	19.90	19.90	12	15.40	21.73	21.14
3	29	16.05	23.13	21.00	10	14.28	21.08	20.81	5	13.21	19.68	19.90	18	16.69	23.02	21.14
4	25	15.70	20.65	21.41	32	14.10	21.40	20.63	17	12.53	18.86	19.78	42	14.61	19.92	21.56
5	7	13.79	20.30	21.37	41	13.11	20.06	20.69	37	14.06	20.31	19.77	27	15.64	20.44	21.10
6	21	15.63	21.94	21.34	3	12.87	20.28	20.57	46	14.80	20.17	19.17	28	16.69	22.69	20.76
7	38	15.31	20.83	21.47	20	13.39	20.61	20.40	26	15.25	17.21	19.02	30	13.76	19.41	21.10
8	9	16.32	22.48	21.74	13	13.42	20.52	20.57	6	14.71	18.70	18.43	33	15.88	21.32	20.93
9	43	15.91	21.81	21.60	15	13.44	20.54	20.65	39	12.67	18.12	18.09	40	15.03	21.62	20.33
10	47	16.41	21.65	21.33	2	13.92	21.39	20.68	14	12.29	17.37	18.30	22	12.66	19.60	20.21
11	16	15.28	21.25	21.06	23	12.33	20.20	20.06	4	13.45	18.45	17.93	44	15.82	19.71	20.26
12	48	15.01	19.50	20.67	24	13.01	20.76	19.69	1	12.32	18.80	17.78	45	13.29	18.78	19.57
13	42	15.78	21.09	20.29	36	12.47	17.40	18.88	11	10.81	16.84	18.21	3	14.20	21.61	19.59
14	30	14.22	19.87	20.23	29	11.63	18.71	18.40	18	11.02	17.35	18.17	5	12.17	18.64	19.28
15	33	14.29	19.73	20.39	25	12.39	17.35	17.57	19	13.16	19.54	18.34	10	12.43	19.23	19.41
16	8	15.45	20.94	20.29	34	11.30	17.78	17.72	20	11.04	18.26	18.87	17	11.79	18.12	18.74
17	6	15.31	20.30	20.27	38	11.08	16.60	17.73	41	12.78	19.73	18.85	24	11.70	19.45	18.65
18	46	15.26	20.63	20.66	7	11.63	18.14	17.63	13	12.39	19.49	18.25	31	12.00	18.27	18.47
19	44	13.86	19.79	20.61	9	12.58	18.78	17.55	28	11.21	17.81	18.41	35	11.74	18.16	18.86
20	26	17.10	21.58	20.47	21	10.55	16.86	17.64	12	10.23	16.56	17.75	1	11.77	18.33	18.67
21	45	15.22	20.71	20.34	47	12.15	17.37	17.29	23	11.17	19.04	17.27	32	12.97	20.07	18.51
22	4	14.58	19.58	20.55	43	11.16	17.06	17.11	40	9.87	16.46	17.50	37	12.27	18.53	18.49
23	27	15.19	19.99	20.63	48	11.89	16.38	17.30	22	10.16	17.10	17.91	39	12.00	17.45	18.57
24	14	15.71	20.79	20.69	16	11.91	17.88	17.44	15	11.24	19.34	18.24	2	10.60	18.07	18.76
25	31	15.81	22.08	21.34	30	12.17	17.82	17.49	32	11.33	18.63	18.16	11	13.20	19.23	19.49
26	18	14.67	21.00	21.39	8	13.58	18.07	17.94	10	13.87	20.67	18.48	13	13.43	20.53	19.77
27	5	16.36	22.83	21.44	42	11.99	17.30	17.90	29	8.99	16.07	18.67	15	15.08	22.18	19.70
28	35	13.81	20.23	21.27	33	13.19	18.63	18.12	25	13.73	18.69	18.51	19	12.45	18.33	19.90
29	17	14.75	21.08	21.34	26	13.50	18.08	18.12	21	13.00	19.31	18.32	16	11.75	17.72	20.19
30	37	14.97	21.23	20.82	44	12.62	18.51	18.02	9	11.71	17.81	18.87	20	13.47	20.69	19.95
31	12	14.98	21.31	21.04	6	13.69	18.07	17.77	7	13.21	19.72	18.48	23	13.67	21.54	19.89
32	28	14.26	20.24	21.37	46	11.45	16.82	17.76	3	11.42	18.83	18.55	34	14.51	20.99	20.42
33	22	14.38	21.32	21.24	45	11.86	17.35	17.74	47	11.40	16.62	18.68	36	13.59	18.52	20.52
34	1	16.18	22.74	21.32	27	13.23	18.03	18.03	24	12.00	19.75	18.95	38	14.85	20.37	20.30
35	40	14.00	20.59	21.54	4	13.43	18.43	18.43	43	12.59	18.49	19.22	41	14.21	21.16	20.28
36	39	16.28	21.73	21.82	14	14.43	19.51	18.91	2	13.57	21.04	19.96	48	15.96	20.45	21.24
37	19	14.94	21.32	21.07	5	12.38	18.85	19.03	34	13.71	20.19	20.59	9	14.81	20.91	21.65
38	11	16.68	22.71	20.93	31	13.48	19.75	19.52	33	14.88	20.32	20.97	25	18.35	23.31	21.84
39	10	12.20	19.00	20.67	18	12.30	18.63	19.69	30	17.26	22.91	21.20	26	17.83	22.41	22.33
40	32	12.59	19.89	20.22	35	14.43	20.85	19.91	8	15.90	20.39	21.57	29	15.02	22.10	22.55
41	20	13.20	20.42	19.53	12	14.06	20.39	19.33	36	16.68	22.20	22.34	6	18.55	22.94	22.40
42	41	12.11	19.06	19.62	37	13.65	19.91	20.59	44	16.13	22.02	22.09	21	15.58	21.89	22.28
43	3	11.87	19.28	19.43	28*	13.86	19.86	20.68	35	19.26	24.19	22.64	43	16.74	22.64	22.56
44	13	12.36	19.46	19.84	17	15.59	21.92	20.62	42	16.36	21.67	22.84	7	15.72	21.85	22.44
45	15	11.82	18.92	19.27	40	14.74	21.33	21.03	45	17.65	23.14	22.74	4	18.52	23.52	22.53
46	2	12.01	19.48	19.42	1	13.50	20.06	21.60	16	17.19	23.16	22.63	14	17.23	22.31	22.68
47	23	11.33	19.20	19.42*	22	15.02	21.96	21.60*	27	16.73	21.53	22.63*	46	16.99	22.36	22.88*
48	24	12.29	20.04	19.42*	39	17.26	22.71	21.60*	48	19.18	23.67	22.63*	47	19.15	24.77	22.88*

** Plus 20, to avoid negative values.
* The moving average for third from first and last rows in each series are used also for first and last two rows.

Under the assumption that 48 strains were grown in the experiment, this direct process is not permissible. The diversity to be expected among the strain means would lead to an unduly large standard devia-

tion. At least three ways of eliminating the effect of the diversity in the strain means are available, differing only in the manner of reaching the same end point.

1. The correlation between the individual deviations and the strain mean deviations may be computed, and from this may be obtained the partial variance of the experiment for a constant strain mean value which must be multiplied by 4/3, to correct for the fact that each of the 4 plots of a strain is involved in determining its mean which is used in computing the correlation. The root of this is the partial standard deviation of the experiment for a constant strain mean.

2. The sum of the squares of the deviations of the strain means around the general mean may be multiplied by the number of plots of each, 4, and subtracted from the sum of the squares of the individual deviations around the general mean. The remainder is divided by 144, the number of degrees of freedom obtained by subtracting 47 (one less than the number of varieties) from 191 (one less than the total number of plots). The root of the mean square so obtained is identical with the partial standard deviation obtained under (1).

3. The deviation of each individual plot yield may be taken from the mean of the strain grown in that plot. These deviations, plus 20 pounds to avoid negative values during computations, are shown in columns 4 of Table I, as F. Multiplying the mean square of these deviations by 4/3 and extracting the root gives the same standard deviation as by (1) and (2).

Obviously, to divide by 192 and then multiply by 4/3 under method (3), is the same as to divide by 144 as was done under (2). It seems worth while, however, to note the different steps and the superficially different theories for using 144, or 3/4 of 192. In determining the correlation on which the partial standard deviation of method (1) was based, the product moment included a portion due to the four plot deviations of a strain constituting the strain mean with which they were correlated. Under method (2), the 144 degrees of freedom represent the 144 values that could be filled in at random, after which the other 48 would be determined. In method (3), the four plots of a strain constitute a sample of a larger population, the mean of which would be expected to differ from that of the sample. The standard deviation of the population, which is the one of interest, would be expected to be $\frac{n}{n-1}$, or 4/3 times the standard deviation of the sample.

The writer prefers the third method for experiments of this kind. It seems more simple and direct. It may be designated the deviation

from the mean method as modified by "Student."[1] In some of Fisher's [2] arrangements, however, where different treatments are the variables, method (2), Fisher's analysis of variance, is the most simple. Whichever method is used, the final value is the same. This value may be designated σ_F, the standard deviation in the "Field," apart from the influence of strain means differing inherently among themselves and therefore more than would be expected under random sampling.

In the example under consideration the means of the assumed strains should be the same within the limits of random sampling. Accordingly, σ_F here should approximate σ_T (=1.999) as determined from all of the plots, in so far as theory and fact are in agreement. Actually, the value of σ_F is 2.053, a discrepancy of only 2.7 per cent. Where the strain or treatment means may be expected to differ, σ_F provides the best available estimate of the gross experimental error in experiments of this kind. It therefore has wide application. In many agronomic experiments, however, it is needlessly large because of the influence of systematic soil variation, thus making the results seem less precise than they really are.

The field variation, measured by σ_F, is caused to a very large extent by differences in soil productivity from one part of the experimental field to another. Some of these are random differences. Others are systematic and tend to affect adjacent plots alike, thereby bringing about a correlation among the individual deviations within certain parts of the field. To the extent that this correlated variation may be measured and its effect eliminated, the precision of the estimate of the experimental error is increased. Various methods have been used in eliminating this correlated variation. Many of these have been reviewed by Hayes and Garber,[3] and by Stadler,[4] and it seems better to restrict the discussion here to certain specific methods.

In 1923, "Student" [5] suggested a method for estimating the experimental error with the effect of variation from replication to replication eliminated, giving credit to R. A. Fisher for the final solution of the problem. The effect of this simple method was the same as to determine the differences between all strains in all combinations within each replication and compute the mean standard deviation for all differences. It already has been pointed out [6] that this is equivalent to determining

[1] H. K. Hayes and R. J. Garber, *Breeding Crop Plants*, New York, McGraw Hill, 2nd Ed., 2nd impression, 1927, p. 84.

[2] R. A. Fisher, *Statistical Methods for Research Workers*, London, Oliver and Boyd, 2nd Ed., 1928.

[3] Hayes and Garber, *op. cit.*

[4] L. J. Stadler, "Experimental Error in Field Plot Tests," *International Congress of Plant Sciences, Ithaca, 1926, Proceedings*, 1: 107–127, 1929.

[5] "Student," "On Testing Varieties of Cereals," *Biometrika*, 15: 271–293, 1923.

[6] F. D. Richey, "The Moving Average as a Basis for Measuring Correlated Variation in Agronomic Experiments," *Journal of Agricultural Research*, 32: 1161–1175, 1926.

the partial average standard deviation of all strains for a constant replication value and multiplying this by $\sqrt{\frac{m}{m-1}}$, where m is the number of strains. The principle involved is entirely similar to that on the basis of which the effect of differences in varietal means was eliminated. The method really comprises the first step in Fisher's analysis of variance method and is discussed in connection therewith.

Fisher's analysis of variance method for eliminating correlated variation due to soil differences in two directions is shown in its application to the uniform-planting experiment of Table I, in Table II. The sum

TABLE II

ANALYSIS OF VARIANCE BY R. A. FISHER'S METHOD

Variation due to	Degrees of freedom	Sum of squares	Symbol	Variance σ^2	σ
Total	191	763.5940	...	3.9979	1.999
Variety	47	156.7964	...	.....	.....
Experiment and position	144	606.7976	σ_p	4.2139	2.053
Columns (series)	3	83.8057	...	.....	.....
Experiment and groups	141	522.9919	${}_G\sigma_p$	3.7092	1.926
12-row groups	3	176.0951	...	.....	.....
Experiment	138	346.8968	${}_P\sigma_p$	2.5137	1.585

of the squares of the deviations of the individual plots around the mean of all plots, 763+, constitutes the base. From this is subtracted 4 times the sum of the squares of the varietal means around the general mean, or 156+. The difference, 606+, is equal to the sum of the deviations of the plot yields from their respective varietal means, i.e., the sum of the squares of F deviations. This is reduced by subtracting the variation that is common to all the strains from series to series, or 48 times the sum of the squared deviations of the series means around the general mean, giving the sum of the squares for "Student's" equivalent of the "Pairing" method, under which designation its standard deviation is shown as ${}_G\sigma_F$ in Table IV. This sum finally is decreased by eliminating the variation common to the 12-row groups across the series. The two successive reductions are shown in Table II, together with the number of "degrees of freedom" by which each sum of squares should be divided to obtain an estimate of the variance due to the different causes. The standard deviations obtained under these different methods will be discussed later in comparison with that which is obtained by the moving average method.

In an earlier paper [1] it was suggested that a moving average of the F deviations might be used as an Index of Productivity (I) to estimate correlated variation in experiments of this kind. Using a 5-plot moving average (I_5), and with a different (random) order in the four series, the mean of the deviations of any 5 adjacent plots affords an excellent estimate of the inherent productivity of the soil on which these plots, and more specifically the central plot of the 5, occurred. The I_5 values for the present experiment are shown in columns 5 of Table I, again with 20 pounds added to avoid negative values. Each moving average value is recorded for the plot on which the 5 plots are centered, except for rows (plots) 1 and 2, and 47 and 48 in each series for which the I values for plots 3 and 46 are used. The method suggested was to determine the correlation between the F and the I deviations, the regression of F on I, and the partial standard deviation of F for constant I. The latter, of course, would be the standard error of estimate remaining after the F values had been adjusted to their regression on the corresponding I values, but could be had directly from the formula:

$$_I\sigma^2_F = \sigma^2_F \ (1 - r^2_{FI}).$$

Multiplying this partial variance by 5/4, to correct for the fact that the central plot deviations were used in obtaining the averages with which they were correlated, and extracting the square root gave the standard error for the experiment, freed from the effects of such correlated variation as was measured by the moving average. The results of using this method with the data of Table I are shown in Table III under the 5-row moving average and in Table IV as the approximate moving average method.

TABLE III

ADJUSTMENT FOR REGRESSION ON A MOVING AVERAGE

Kind of moving average	r_{FI}	$\text{regr}_{F.I}$	Gross σ^2_F	$(1-r^2_{FI})$	$_I\sigma^2_F$	$_I\sigma_F$
4-rows (2 on each side)	0.7596	0.8969	4.2139	0.4230	1.7825	1.335
5-rows centered	.8512	1.0138	4.2139	.2755	1.4511	1.205

The multiplication of the partial variance for a constant moving average value by 5/4 to correct for including the central plot must be regarded as an approximation. The direct way would be to use a 4-plot moving average based on the 2 plots to each side of a central plot. Computing such a 4-plot moving average is exceedingly la-

[1] Richey, *op. cit.*

borious, whereas computing the 5-plot average is simple. It was for this reason that the method described was followed. More recently a simple means was found for converting correlation and regression values determined from the 5-plot average to a 4-plot basis, with no discrepancy other than that entering from the fact that there are no true moving averages for the plots at the ends of the series.

Symbolizing a moving average of 4 rows, 2 on each side of the row on which the moving average is centered, as I_4, the formulas for obtaining the mean product moment and observed standard deviation on a 4-plot basis (p_{FI_4} and $S^2_{I_4}$) from the corresponding values on a 5-plot basis (p_{FI_5} and $S^2_{I_5}$) and from the mean square of the F values (S^2_F) are as follows:

$$p_{FI_4}=\tfrac{1}{4}(5p_{FI_5}-S^2_F),$$

or, in the example, $$=\tfrac{1}{4}(5\times 2.2586-3.1604)=2.0332$$

$$S^2_{I_4}=1/16(25S^2_{I_5}-10p_{FI_5}+S^2_F)$$

or, in the example, $$=1/16(25\times 2.279-10\times 2.2586+2.1604)=2.2670.$$

The correlation between the F deviations and a 4-row moving average is 0.7596; $1-r^2_{FI_4}=0.4230$, and the product of these two, ${}_{I_4}\sigma^2_F$, is 1.7825, as shown in Table III. This is 58 per cent less than the total variance in F, the reduction being due to the elimination of correlated variation measured by the 4-row moving average. The square root of this is shown in Table IV as for the moving average method.

TABLE IV

COMPARISON OF STANDARD DEVIATIONS OBTAINED BY DIFFERENT METHODS

Designation of method	Symbol *	Value	Percentage of σ_F	Relative variance
Deviation from the mean	σ_F	2.053	100	100
Pairing	${}_G\sigma_F$	1.926	94	88
Analysis of variance	${}_P\sigma_F$	1.585	77	60
Moving average	${}_{I_5}\sigma_F$	1.335	65	42
Moving average, approximate †	${}_{I_4}\sigma_F$	1.205	59	34

* As used here. There is no generality of usage.
† Shown for completeness only.

The estimates of the experimental error obtained under the different methods are brought together for comparison in Table IV. The method which eliminates only the effect of the variation from replication to replication is indicated as the pairing method, reserving the term, analysis of variance, for the method which eliminates the effects

of variation in both directions. The approximate moving average method is included only for completeness, as its difference from the method based on a 4-row moving average is due only to the approximation involved.

Beginning with the deviation from the mean method, the standard deviation of F $(=2.053)$ represents the best estimate of the standard deviation in the experiment due to all causes other than strain differences. The standard deviations obtained by the more refined methods are smaller because definitely measurable parts of the correlated variation were eliminated. Thus, the standard deviation has been reduced 6 per cent by eliminating the effect of variation from series to series. This can be effected in nearly all replicated experiments, although if the number of items being compared is large and the number of replications small, the reduction may be unimportant. Eliminating the effect of variation in the 12-row groups across the series effects a further reduction of 17 per cent, or 23 per cent in all. To achieve this, the experiment must have been planned to permit it. Finally, under the moving average method, the standard deviation is reduced 35 per cent. This method requires only that the strains shall have been planted in a different or random order in the different replications. Falsely high correlations may be obtained if the same order of planting is used in each replication.

Which method will be best in any experiment will depend largely upon the experiment itself. With relatively few items to be compared, the area required for each replication is small, and the analysis of variance method may be quite as effective as the moving average method in obtaining a small standard deviation. With a large number of items to be compared, the moving average method usually will provide a more adequate basis for eliminating correlated variation. Even so, the labor of computing the moving average may be excessive in comparison with the increase in the precision of measuring the probable significance of differences. The labor of the moving average method is not unduly serious, however, if computing-tabulating machinery is available and the methods suggested by Snedecor [1] and Brandt [2] are used.

There is a second advantage in using the moving average method. The final estimate of the standard deviation $_1\sigma_F$ is the standard error of estimate that would remain after adjusting the individual yields to their regression on the moving average. The mean of the moving

[1] G. W. Snedecor, "Uses of Punched Card Equipment in Mathematics," *American Mathematical Monthly*, 35: 161–169, 1928.

[2] A. E. Brandt, "The Use of Machine Factoring in Multiple Correlation," this JOURNAL, September, 1928.

average values for the plots of any strain furnishes a valuable estimate of the average productivity of the soil in which the plots of that strain were grown. Even though the 5-plot moving average values are used to obtain this mean, the latter is unaffected by the inclusion of the central rows, the sum of the deviations of which is zero. Such a mean moving average value, however, will be only 4/5 as large as had the 4-row moving average been used. This may be compensated for by multiplying the regression of F on I_4 by 5/4, and multiplying the mean I_5 values for the different strains by the value so obtained. In the example, the regression of F on I_4 is 0.8969, which, multiplied by 5/4, gives 1.1211, the factor to use with the mean deviation of I_5.

The mean yields of the 48 assumed strains are shown in column 2 of Table V. The deviation of each strain mean from the general mean is shown in column 3. The mean of the four 5-plot moving averages for each strain (after deducting the 20 pounds previously added for coding) multiplied by 1.1211, are shown in column 4, with the signs changed so that they represent the necessary adjustment for regression on the moving average. Finally, the deviations of the strain means after adjustment from the general mean are shown in column 5. The deviations of the strain means from the general mean are less after adjustment for 34 of the 48 strains, the standard deviations being 0.91 before adjustment and 0.72 after adjustment, a reduction of 21 per cent. When it is remembered that the 48 "strains" were really the same kind of corn in this case, and therefore should have yielded equally, it would seem that a real gain in the precision of measuring their productiveness had been achieved.

Perhaps the moving average method has its greatest usefulness in breeding experiments where many selections must be tested with the primary purpose of determining the few best. Thus, in the coöperative corn breeding experiments at Ames, Iowa, from 400 to 600 different crosses between inbred lines of corn are being tested each year. Such an experiment covers 10 to 20 acres, each cross being grown in six 13- to 15-hill plots distributed at random within each of 6 series. The regression on a moving average has been of much use in adjusting for soil heterogeneity in these experiments.

In plant breeding experiments it frequently is desirable to measure the extent to which several characters are interrelated or are associated with some other character. Thus, in selecting corn for productiveness, will strains having long ears be more productive than those having ears of large diameter, and is there any general relation between length and diameter of ear? Again, is resistance to some specific disease, or erectness of plants, their height, leafiness, and the like, related to yield in the

TABLE V

THE MEAN YIELDS OF THE 48 ASSUMED STRAINS OF CORN AND THEIR DEVIATIONS FROM THE GENERAL MEAN OF THE EXPERIMENT BEFORE AND AFTER ADJUSTING TO THEIR REGRESSION ON A 4-ROW MOVING AVERAGE

Strain	Strain mean	Deviation from general mean	Adjustment for moving average	Deviation after adjustment	Strain	Strain mean	Deviation from general mean	Adjustment for moving average	Deviation after adjustment
1	2	3	4	5	1	2	3	4	5
1	13.44	−0.47	0.18	−0.29	25	15.04	1.13	.19	1.32
2	12.53	−1.38	.33	−1.05	26	15.42	1.51	.01	1.52
3	12.59	−1.32	.52	−.80	27	15.20	1.29	−.67	.62
4	15.00	1.09	.16	1.25	28	14.00	.09	−.35	−.26
5	13.53	−.38	.10	−.28	29	12.92	−.99	−.17	−1.16
6	15.61	1.70	.31	2.01	30	14.35	.44	.01	.43
7	13.49	−.42	.02	−.40	31	13.73	−.18	.21	.03
8	15.51	1.60	−.27	1.33	32	12.70	−1.21	.70	−.51
9	13.90	−.01	.06	.05	33	14.56	.65	−.11	.54
10	13.20	−.71	.18	−.53	34	13.52	−.39	.08	−.31
11	13.97	.06	.16	.22	35	13.58	−.33	.01	−.32
12	13.67	−.24	.03	−.21	36	15.07	1.16	−.85	.31
13	12.90	−1.01	.58	−.43	37	13.74	−.17	.20	.03
14	14.92	1.01	−.22	.79	38	14.48	.57	−.52	.05
15	12.90	−1.01	.59	−.42	39	14.55	.64	−.06	.58
16	14.03	.12	−.37	−.25	40	13.41	−.50	−.11	−.61
17	13.67	−.24	−.13	−.37	41	13.05	−.86	.16	−.70
18	13.67	−.24	−.11	−.35	42	14.69	.78	−.75	.03
19	13.62	−.29	−.03	−.32	43	14.10	.19	−.13	.06
20	12.78	−1.13	.35	−.78	44	14.11	.20	−.28	−.08
21	13.69	−.22	.15	−.07	45	14.51	.60	−.13	.47
22	13.06	−.85	−.27	−1.12	46	14.63	.72	−.13	.59
23	12.13	−1.78	.94	−.84	47	14.78	.87	−.06	.81
24	12.25	−1.66	.92	−.74	48	15.51	1.60	−.52	1.08

given environment? The methods of simple, partial, and multiple correlation, together with the corresponding regressions, are valuable tools in such experiments and are in wide use.

A recent paper on "Correlation Studies with Inbred and Crossbred Strains of Maize" [1] contains numerous examples of such uses of correlations. Its detailed consideration here is unwarranted because of its general availability. It is worth noting, however, that, in addition to a study of relations of various characters within a single generation, the correlations also are presented for characters in the inbred parent strains with the same or different characters in the crossbred progeny. Thus, a coefficient of 0.43 ± 0.03 for length of ear in parent inbreds and their crossbred progenies affords an estimate of the extent to which variation in length of ear in the parent strains was reflected in the crosses. Similarly, a coefficient of correlation of 0.20 ± 0.03 between the yield of the inbred parents and that of their crossbred progenies indicates that the higher yielding inbreds tended, as an average, to produce the higher yielding crosses in this experiment, a condition that might or might not be expected from different theoretical considerations.

The importance of a correlation of 0.20 may be questioned by some. It must be remembered, however, that these coefficients are based on data from field experiments involving a large amount of random soil variation. They, therefore, are relatively much more important than their size indicates. Furthermore, selection toward a uniform type of parent tends to decrease parent-progeny correlations at the same time that fixation of type otherwise would tend to increase these correlations.

These last conflicting tendencies may be shown with data from coöperative breeding experiments at Columbus, Ohio. The time of silking in corn is affected both by environment and heredity, and uniformity in the time of silking is highly desirable in sweet corn for canning. The object was to eliminate as much as possible of the variation in time of silking due to heredity. The data for four years are shown in Table VI. The period during which the original self pollinations were made in 1924, was 15 days. This period was decreased in the later generations, being only 6 days in 1927. The standard deviation for the mean time of silking of the progenies decreased from 3.7 in 1925, to 2.3 in 1926 and 1927, and then increased to 4.5 in 1928. Apparently, then, there was little advantage from the selection for uniformity in time of silking. Column 5 of Table VI, however, shows that the coefficient of correlation between the date of pollinating a

[1] M. T. Jenkins, in the *Journal of Agricultural Research*, 39: 677–721, 1929.

parent ear (practically a constant function of date of silking of the plant on which it was produced) and the mean date of silking of the progeny grown from seed of that ear, decreased progressively from 0.75 in 1925 to 0.08 in 1928. This simply means that the variation among the parent ears pollinated in the later years was due to environment rather than heredity and, accordingly, was not reflected in their progenies.

TABLE VI

MEASUREMENT OF THE EFFECTS OF SELECTION IN SELF-FERTILIZED LINES OF CORN FOR UNIFORMITY IN DATE OF SILKING AT COLUMBUS, OHIO

(Data from M. T. Meyers and L. R. Jorgenson)

Year	Number of progenies grown	Number of days within which parent plants were pollinated	σ for number of days from planting to silking among progeny means	Correlation between date of pollinating parent ear and mean number of days to silking in progeny
1	2	3	4	5
1924	..	15	.....	.....
1925	22	9	3.7 ± 0.37	0.75 ± 0.056 *
1926	36	7	2.3 ± .18	.68 ± .060
1927	47	6	2.3 ± .16	.40 ± .084
1928	72	..	4.5 ± .36	.08 ± .079

* Only 21 progenies used in correlation study.

A similar situation occurs when selection for any characteristic is successful to an extent that eliminates hereditary differences. If, on the other hand, selection had been directed to obtaining strains which silked at different times, the fact that hereditary variation had been largely eliminated from each strain while the strains differed hereditarily among themselves, would have resulted in a very high parent progeny correlation in the population as a whole. The conflicting tendencies of these two factors must always be kept in mind in interpreting parent progeny correlations.

The results of the 1928 experiments on breeding corn resistant to the European corn borer, at Bono, Ohio, coöperative with the agronomists and entomologists of the Ohio Agricultural Experiment Station, have been drawn on for another example of the application of statistical methods. It first may be stated briefly that the corn borer moths emerge during a period of some 30 days centered more or less about July 10 to 15, when the peak of the moth flight occurs. The female moths deposit the eggs in clusters of about 15, mostly on the surface of the corn leaves. The young corn borer larvae feed on the leaves until they finally bore into some part of the stalk. They are extremely active and there is a heavy mortality so that a relatively small percentage persist to reach a stage and a place where they do damage.

The experiments in 1927 had shown that the percentage of plants infested in the different plots was correlated negatively with the date on which the plants in these plots had silked, and positively with the height of the plants during the period of egg deposition. That is, the earlier silking plots and the plots which were taller about July 10 tended to have the larger percentages of stalks infested with borers at harvest. Accordingly, the experiments in 1928 were planned to get more definite information on this relation and, particularly, to differentiate between possible "resistance" caused by a lack of attractiveness resulting in fewer eggs being deposited on some plants, and that which might be caused by a larger mortality from the egg to the final larval stage. The coefficients of correlation in Table VII show the results of such a study in a single experiment in which 25 different strains were grown in duplicate plots.

TABLE VII

COEFFICIENTS OF SIMPLE CORRELATION AMONG THE VARIABLES SHOWN BASED ON DATA FROM THE DUPLICATE PLOTS OF EXPERIMENT A AT BONO, OHIO, IN 1928

(Data from M. T. Meyers and L. L. Huber)

Variable	Symbol	y	d	$h4$	$h6$	e	p	s
Yield of corn	y							
Days to silking	d	−0.23						
Plant height, July 20	$h4$	.32	− .35					
Mature plant height	$h6$	.09	.79	.11				
Corn borer eggs per plot	e	.25	− .43	.67	.00			
Corn borer larvae per plot	p	.07	− .58	.54	−.28	.44		
Percentage of survival	s	− .06	− .30	−.03	−.34	...	...	
Percentage of stalks infested	i	.09	− .75	.66	−.45	.60	...	...

The correlations indicate that the earlier silking strains tended to yield slightly more than the later ones in this experiment and to be somewhat taller on July 20, though materially shorter at maturity. More eggs were deposited on the earlier silking strains and a larger percentage of the larvae survived through the various developmental stages until harvest. This is reflected in the larger final populations per plot and the larger percentages of infested stalks among the earlier silking sorts, as indicated by the coefficients. The plots which were taller on July 20 tended even more strongly than the early silking ones to become egg-infested. Plant height on July 20, however, was uncorrelated with survival and the relations for plant height, July 20, with final population and with percentage of stalks infested are no more important than the corresponding ones for time of silking. The only other relations meriting comment here are those for corn borer eggs per plot with final population (0.44) and with percentage of stalks

infested (0.60). Only some 19 per cent of the variance in population and 36 per cent of that in infestation is determined by the variation in numbers of eggs deposited. This is encouraging from the standpoint of the program. Resistance, if any, that was due to relative unattractiveness easily might be lost if all of the area were of the less attractive sort. Failure to survive, on the other hand, would not be influenced by having the entire area planted to the same kind of corn.

The important influence of time of silking and early plant height on infestation requires that these factors be taken into definite account before concluding that a strain tends to be more resistant or susceptible than the average. The multiple regression equation for borer population on time of silking and early plant height provides a convenient tool through which this can be accomplished. The population to be expected in each strain may be predicted from its time of silking and early plant height. The deviations from these predicted values then can be studied instead of the deviations from the mean.

Of the strains in experiments A, D, and E at Bono, Ohio, in 1928, 14 were tested in each of the three experiments and another 19 were in experiments D and E, but not in experiment A. The deviations of the corn borer populations observed in these strains from the populations predicted from the multiple regression on time of silking and early plant height are shown in Table VIII. Experiment A consisted of duplicate plots planted on May 22, experiment E consisted of single plots planted on May 23, and experiment D consisted of duplicate plots planted on June 8, the later planting tending to reduce the extent of infestation.

TABLE VIII

DEVIATIONS OF THE OBSERVED CORN BORER POPULATIONS FROM THE POPULATIONS PREDICTED ON THE BASIS OF THE MULTIPLE REGRESSION OF POPULATION ON DATE OF SILKING AND EARLY PLANT HEIGHT IN 33 STRAINS OF CORN GROWN IN THREE EXPERIMENTS AT BONO, OHIO, IN 1928

(Data from M. T. Meyers and L. L. Huber)

Deviations from predicted in experiments								
D	E	A	D	E	A	D	E	A
−1.1	−21.3	−23.8	5.3	3.6	7.5	− .5	2.1	−9.5
−2.4	−22.0	−20.8	2.4	5.8	....	−1.3	7.7	...
−7.7	−22.9	−18.3	2.0	8.4	....	4.6	−25.9	...
−7.2	−27.0	−25.7	2.7	18.1	....	2.8	−13.7	...
−1.9	− 4.1	− 5.7	14.9	45.1	....	7.9	− 6.3	...
−2.1	− 8.1	....	−00.8	−08.5	4.9	−3.3	5.5	...
−5.4	− .3	....	− .2	− 4.5	11.4	−4.0	15.4	...
−4.6	−16.8	....	− 5.8	− 8.4	14.4	−2.8	48.9	...
−5.7	−27.1	....	2.4	11.2	− 7.3	−1.4	16.8	...
− .7	− 7.5	....	2.3	16.7	− 5.8	1.4	−7.8	...
3.4	12.5	72.0	− 1.9	5.4	−15.3	−2.7	5.8	...

Of the 14 strains that were in all three experiments, the deviations of 7 are alike in sign throughout, whereas those of the 7 other strains differ between experiments. If the signs of the deviations depended upon chance alone they would be expected to agree only in ¼, or 3.5 of the strains, disagreeing in the other ¾. The odds are 25 to 1 against 7 or more agreements in 14 trials as a result of chance. Similarly, of the 33 strains in experiments D and E, the deviations for 21 agree in sign and those for 12 differ in the two experiments, a departure from the equality to be expected under random sampling of 4.5. The odds against chance occurrence of 21 agreements in 33 trials are 11 to 1.[1] Although far from conclusive, the results indicate that there are inherent differences in resistance and susceptibility among the strains. Methods taking account of the size as well as the direction of these deviations also may be used to estimate their probable significance, but need not be considered here.

The foregoing exemplify some applications of statistical methods to comparisons of hereditarily different plants in the same general environment. The same principles are being utilized in experiments to determine the effect of definite environmental differences on plants of a single strain or the differential effects on plants of several strains. Regardless of the intentional variation involved in the experiment, however, there always is an additional element of variation because of the inability to obtain entirely uniform plants or constant environments in which to grow them. The data from an agronomic experiment, therefore, constitute the records of a more or less complex system of variables. The concepts and applications of statistical methods have been more and more useful to agronomists as they have become better understood and, particularly, as agronomic experiments have increased in complexity.

[1] The writer's attention was called by Professor Harold Hotelling to the fact that the probable errors and odds originally presented here and based on the normal distribution with $\sigma = \sqrt{n\,p\,q}$ were unreliable because of the small values of n. Calculation from the expanded binomials, $(¼ + ¾)^{14}$ and $(½ + ½)^{33}$, gives the above odds against the chance occurrence of 7 or more and 21 or more agreements in the two comparisons.

SOURCES OF INCOMPARABILITY AND ERROR IN EMPLOYMENT-UNEMPLOYMENT SURVEYS

By Margaret H. Hogg, *Russell Sage Foundation*

Interest in employment-unemployment surveys has increased rapidly of late, but the number of surveys made and the public attention directed to their results are not more striking than the diversity of investigation-procedure adopted. Comparison of past surveys proves that variation in procedure can by itself produce very great difference in the degree of unemployment found; it seems, therefore, very desirable that some standard procedure be adopted for general use. In some details variation may be inevitable, but much past variation has been optional; such surveys will never attain their full value until the methods used are not only sound but also productive of comparable results. In the hope of hastening agreement on a desirable procedure, some important sources of variation and of error in recent surveys are here discussed.

Since unemployment and idleness rates are the first objective of such surveys, uniform definitions of the unemployed person and of the person liable to unemployment are essential to comparability, for the total of unemployed persons must be the numerator and the total of persons liable to unemployment must be the denominator of the unemployment rate. It is clear that employers, persons working on their own account, persons permanently incapacitated, and other persons not usually employed, are none of them liable to unemployment. Therefore, none of them should be admitted to the total on which unemployment rates are based. Yet all of these have sometimes been included when making so-called unemployment rates, greatly increasing the resulting figures. Obviously, satisfactory tests for employment-unemployment categories are very difficult to devise and to apply, yet this is no adequate reason for throwing in all permanent stay-at-homes with the temporarily idle. Suitable tests are immensely important, and unless the same tests are used in different surveys the effect on comparisons may be quite upsetting. Is a man to be regarded as looking for a job if he goes to an employment bureau once a week, or more often, or must he spend his days in searching? Decision on this point will have great influence on the numbers regarded as unemployed. Since the tests must be applied in conversational interviews by teams of investigators who have very varied backgrounds, it is not sufficient that a test should be logically suitable; it must also be easily

applicable. Appropriate tests should surely be evolved by experts, and published with very careful explanation. With the support of an expert publication, survey-makers would probably have lessened difficulty in getting suitable tests applied.

ENFORCED RETIREMENT

Of all distinctions for which tests have to be devised, probably none presents more difficulty than the essential distinction between the temporarily unemployed and the person regarded by employers as no longer eligible. There is known to be a lamentable number of persons who desire employment and are apparently capable of work of some description, but who have given up the search for work because of continued failure to obtain it, that is, those who have succumbed to enforced retirement. Whether study of forced retirement is or is not intended, and however difficult it may be to devise and apply subjective tests, these shelved persons who have been forced to accept a condition of permanent withdrawal must be identified. No study can be sound which does not separate these unfortunates, according to some definition, from the ordinary unemployed or otherwise idle workers, and also from the normally employed. The economic significance of the shelved persons is different from that of persons usually employed but temporarily idle; the two groups must be separated before either problem can be dealt with effectively; unless they are kept apart we have no means of disentangling the effect of fluctuation in the demand of industry for labor from the effect of that modification of the tasks set to labor which appears to be rapidly reducing opportunity for the older worker, and we therefore cannot gauge either one. If persons in enforced retirement are included undistinguishably with persons still in the labor market, such procedure must lead at best to confessed failure to identify economic unemployment, at worst it leads to unacknowledged misrepresentation.

There are two good reasons why persons in enforced retirement should be enumerated as well as identified for exclusion. One is that there can be no objective test of employability which will be uniform over periods of years, hence it is most desirable for comparisons between years that a count should always be made of those who have been shouldered out of the employment field against their will by the increasing discrimination against even middle age. The second is that by showing the magnitude of the problem at the particular time, the importance of remedial measures may be demonstrated. Stabilization measures do not solve this problem; the initial necessity is an intensive study first of those capacities which are unimpaired or im-

proved by middle age and second of processes and occupations involving chiefly those capacities. Industry will have to face this issue sooner or later, it cannot go on forever shelving workers at earlier and earlier ages. The sequel of such a practice must be either an unimaginable burden of taxation to cover industrial insurance or pensions or relief, or else unimaginably high wages for the short industrial span to cover the expenses of the rest of life. By study of this category of would-be workers, the public may be the sooner convinced of the importance of this matter.

For satisfactory treatment of the subject of enforced retirement, tabulation is needed giving an age-classification of:

A. Total persons usually employed.
   1. Persons at work.
   2. Persons unemployed (idle from lack of a job).
   3. Persons temporarily idle from sickness or injury.
   4. Persons temporarily idle from other causes.

B. Persons in enforced retirement.

Class A is subdivided into A1, A2, A3, and A4, which should be given also as percentages of class A, the usual base. Class B should also be given as a percentage of class A although not included in it. More subdivisions could profitably be made, but these are essential. In such a table, one would expect to see, with advancing age, first a proportionate rise in class A2, and then an increasing proportion passing out into class B.

Enforced retirement has not been dealt with, in recent surveys, at least. In the Philadelphia survey,[1] those persons for whom superannuation was the cause assigned for idleness may well belong under the heading enforced retirement, although those who stated that they had definitely retired were not enumerated. Removal of the superannuated would reduce the Philadelphia general idleness rate from 10.4 to 9.9 per cent.

The Buffalo survey,[2] like its precursors in Columbus, includes inextricably with the class of men idle though usually employed, not only men in forced retirement, but also those incapable of work and any who have voluntarily retired. This makes it impossible to determine the frequency either of forced retirement or of idleness of persons still in the employment market. Nor indeed can one determine their combined frequency. An unknown number of persons is included

[1] *Monthly Labor Review*, February, 1930, p. 17, "Unemployment Survey of Philadelphia, 1929." Study prepared by J. Frederick Dewhurst and Ernest A. Tupper.

[2] *Unemployment in Buffalo, November, 1929*, by Fred C. Croxton and Frederick E. Croxton, Special Bulletin No. 163 of State of New York, Department of Labor.

who are completely disabled, and whom, therefore, no improvement of employment could affect; nor is the modern world in such a parlous state that one can expect to find no persons alive who have voluntarily retired from work. There exists no way of identifying these irrelevant groups, but a reasonably conservative estimate based on the published tables would put the percentages for men who are idle from all causes at 8.4 of *those usually employed* instead of 10.9 of the *total enumerated* as given in the report, and the true percentage may be even smaller. Idleness rates are always numerically small when on a percentage base, but a difference of 2.5 in the general idleness rate for the city of Buffalo in 1929 would be proportionately considerable, and would represent a difference of over 4,000 in the number of idle men in the labor market. Notice, however, that one cannot guess how many of these 4,000, or more or less, who are not really in the labor market, are the victims of enforced retirement. Evidently if surveys are to give clear information, some subjective tests must be applied which will deal, even if imperfectly, with irrelevant groups.

CLASSIFICATION BY AGE

In some inquiries record of age has not been considered feasible, but even if tactical considerations should sometimes unfortunately preclude a direct inquiry, some indirect questions might perhaps be devised which should provide a rough estimate. Rough grouping in the grades under 25 years, 25 to 44 years, 45 to 64 years, and over 65 years, would be vastly better than nothing, although 10-year grades are highly desirable.

Two facts obviously require that age-grades should start at the 5's. First, the Federal Census grades ages from the 5's in connection with occupation, and comparison with the Census is essential. Second, the notorious heaping of age-statements at round numbers, which disturbs census figures, will certainly be very much greater in a local survey which has neither the prestige nor the legal sanction of the Federal Census; and the effect of mis-statement is greatest if grading is from the 10's.

DURATION OF IDLENESS

The desirability of including a question on duration of idleness is generally recognized. For a question so important, and so often asked, agreement should surely be sought on the scale of time to be used, whether weeks as in the Buffalo survey or months as in the Philadelphia one. Presumably, that unit would be more desirable which would call forth the readier reply from the persons interviewed. If

days should be the unit, it would be necessary to specify clearly whether the number of days is number of days lost from work or number of days elapsed since last day worked. A troublesome problem is that of periods of idleness broken by odd-jobs in a makeshift occupation. Some surveys have asked the time elapsed since the last regular job; this is probably at once the most significant fact and the most easily procured, but if this procedure is adopted the definition of duration must be prominent in publication of the results.

MEANING OF THE TERM "UNEMPLOYED"

The words "unemployed" and "idle" have often been used interchangeably to describe any person who is without work though usually employed, causing waste of one good term. By consistently observing a distinction in all reports, statisticians could exert considerable influence. "Idle" may well be used to denote any person without work though usually employed. "Unemployed" should be reserved to describe persons who are idle because of lack of work, that is, because of the state of the labor market. For the public is certainly apt to consider anything published as an "unemployment rate" as directly measuring the deficit of demand from supply in the labor market. Even a general "idleness rate" is likely to be misconstrued in the same way unless an unemployment rate in the strict sense be given as well, and by this name.

PROCEDURE WHERE A FACT IS NOT REPORTED

Reports have sometimes been published in which tables for the same group have different totals one from another, because persons not reporting the particular fact have been excluded from the body of a table and mentioned in titles or footnotes. Such procedure is bound either to mislead or to annoy, and impairs both comparability between surveys and separate value of a single survey. In a tabular description of a group, all members of it should be accounted for *within the table* by including a category "not reported." Moreover, that category should be included in percentages, and if percentages excluding them are ventured on, the range of possible effect of "not reported" cases should be adequately discussed. If their number is large, the percentages for other categories cannot be assumed to be significant.

RATES OF IDLENESS VERSUS OTHER IDLENESS PERCENTAGES

For some purposes, idleness is naturally studied from absolute numbers of persons idle, but an important function of employment-unemployment surveys is to provide comparisons between dates and

between communities, and for this purpose *rates* of unemployment and idleness are needed.

In comparisons of general idleness, procedure is obvious, but caution is required when general idleness is subdivided by cause, by duration, by industry, or in other ways. *Rates* of idleness are still the first necessity but are sometimes not computed because the fundamental principle has been lost sight of, that rates of contingency must be based on the population at risk. This elementary statistical principle is sometimes transgressed by using persons idle, instead of persons usually employed, as the base for percentage idleness from various causes or of different durations. This has disastrous effects on comparability. The sole necessity for such percentages is where the idle are to be compared with some section of the population resulting from a risk (or a combination of risks) whose incidence is not known.[1] Their only safe use is for comparison of the distribution of the idle with that of all persons usually employed or of some other population group, when the figures are for the *same date* and the *same place*. Even here comparison merely indicates very roughly and imperfectly what *rates* will indicate accurately.

Inferior significance for this simplest comparison already shows that these artificial percentages are unfortunate. But they are worse than that, for other comparisons are sure to be attempted in which they would always mislead. The organic connection of any section of the idle is with the corresponding section of the usually employed. The connection with other groups of idle persons is indirect, and is a product of many factors which cannot be disentangled in percentages based on persons idle. In short, such percentages are artificial, and are too far from the basic relationships to be trustworthy tools for comparison. They are so dangerous in the hands of anyone but an expert statistician that they should be consistently avoided in publication.

In the first place, comparison is generally desired between industries. Take, for instance, the two industries, iron and steel, and building and construction, in Columbus in 1921. The idleness rates are 26.8 per cent for iron and steel, and 13.3 per cent for building and construction, and they contribute 3.1 and 1.0 respectively to the general idleness rate of 13.4 per cent. These are the desirable statements; comparison of the artificial percentages 23.2 and 7.5 based on persons idle gives only a contrast which is equally evident either from the absolute number of persons idle or from the contributions of 3.1 and 1.0, to the general idleness rate. It is so far removed from the basic facts that all indica-

[1] Such a comparison would be that with the case-load of a relief-agency, for which the constitution of the group of persons at risk of having to apply to that relief agency is usually unknown.

tion has been lost both of the prevalence of idleness in either industry and of the general prevalence of idleness, and there is, moreover, serious danger of the false inference being drawn that idleness is more than three times as prevalent in one as in the other.

Next, idleness in the same industry in successive investigations is a likely subject of comparison. Take, for instance, trade and transportation in Columbus in 1921 and 1922. The situation for that group changed very little, the rate of idleness rose only from 7.2 to 7.4, and the contribution of trade and transport to general idleness fell from 2.1 out of the 13.4 per cent in 1921 to 1.8 out of the 7.1 per cent in 1922, the actual number idle falling from 234 to 218. Yet the artificial percentage based on persons idle shows a fantastic rise from 15.8 to 25.3 which is directly due to the exceptional prevalence of idleness in the iron and steel industry in 1921. This is an extreme instance of the artificial nature of such percentages, made expressly for demonstration purposes, but that artificiality is always present in them.

Of comparison of idleness by duration a convenient instance may be found in a table giving figures for 1921 to 1925 in Columbus and for 1929 in Buffalo.[1] Percentages are there based on persons idle instead of on persons enumerated, with the following results if an unwary reader tries to make the comparisons between years which he naturally desires. Comparing successive years only, there are sixteen contrasts, and six of these suggest changes in the reverse direction from the real change. These six reversed contrasts are only the extreme cases of a consistent misrepresentation. Moreover, the statement that of all women in Buffalo in 1929 enumerated as usually occupied, 3.5 per cent had been idle less than ten weeks, and 2.0 per cent had been idle ten weeks or more, would give all the information that could be used from the statement that of women *idle* in Buffalo in 1929, 62.8 per cent had been idle less than ten weeks and 37.2 per cent had been idle ten weeks or more. Further, rates give indication, as the artificial percentages do not, of the real prevalence of idleness of each duration and are comparable with similar rates for Columbus. Any criticism that the true percentages are smaller and have fewer significant figures could be met by carrying the rates to more decimal places or substituting rates per thousand. However, in view of the probable degree of accuracy of this kind of data and of the fact that the investigation covered only a fraction of the households in Buffalo, the accuracy of the artificial percentages is probably quite spurious.

As an instance of comparison of idleness by cause, take idleness of women in 1921 and in 1922 in Columbus due to illness or injury. Here

[1] *Monthly Labor Review*, February, 1930, p. 32.

the suitable statement is that, in 1921, idleness due to illness or injury was 1.9 out of 7.9 per cent general idleness, while in 1922 it was 1.4 out of 3.0 general idleness. If we say that illness and injury contributed 24.7 per cent of idleness in 1921 and 50.0 per cent in 1922, we are obviously straying so far from the basic facts as to present something fantastically artificial; for the percentage of idleness contributed is dominated in this comparison by a sharp drop in the prevalence of idleness due to slack work, which fell from 4.4 per cent to 0.9 per cent. An extreme instance like this, where the figure for idleness due to illness or injury rose to more than double when the prevalence of such idleness actually fell to less than three-quarters, would startle anyone, but misrepresentation which is similar in kind though smaller in degree will lurk in every comparison of the artificial percentages.

The examples adduced should be sufficient to show the general undesirability of percentages based on idle persons. When there is real excuse for such percentages, their very limited usefulness should be specifically indicated. To readers not in need of conversion, this discussion may seem to set up a straw man in order to knock it down again. The justification is that this straw man is set up so astonishingly often with the assumption that it is not an impostor that it evidently needs a good deal of public knocking down.

METHODS OF SAMPLING

The problem of selecting persons or households within the city or region to be studied, while treated last in this discussion, is of paramount importance. The resulting aggregate of persons or households must be demonstrably representative of a population for which an estimate is desired, otherwise the survey will have little or no value.

If a survey is to provide representative figures, rigorous sampling must be maintained. The importance of rigorous sampling has often been overlooked. Sampling by households is probably the most desirable procedure. Selection by blocks has been tried in the Philadelphia survey. It is to be hoped that discussion of this aspect of the survey will be published, for study of the results for individual blocks should be valuable in helping to determine in what circumstances, if any, such a unit of sampling may be used, and what procedure is necessary with it.[1] The safety of block-sampling, however, has not yet been established, and one may doubt whether the saving of effort

[1] It is conceivable that in very large cities stratified sampling by blocks may be justifiable. In other words, proportionate random sampling of blocks may be made by districts (proportionate sampling by districts meaning that each district must have the same proportion between number of blocks in the sample and total number of blocks in the district) even if random sampling by blocks for the city as a whole turns out to be unsafe.

would be commensurate with the lessening of security. The most detailed knowledge humanly obtainable could not provide, unless by the rarest of happy accidents, a properly representative sample by units larger than blocks.

Any published account of a survey needs to show exactly what was done with a view to securing a proper sample, since the results obtained depend directly upon the first process. In other words, in an inquiry by sample it is incumbent upon the makers to demonstrate two things: (1) that the selection was perfectly random; that is, that every block, or every household, had equal chance of inclusion with every other block, or every other household, in the whole city or region being investigated; (2) that according to all tests which are applicable the resulting sample is representative of the whole city.

Both of these conditions are essential. Reassurance by tests is necessary because unsuspected sources of bias may occasionally affect a system of sampling which was supposed to be perfectly random. Reassurance concerning adequate measures to secure a random sample is necessary because a selection may appear representative in some respects and be quite unrepresentative in others; only if a random sample has been secured can one argue from representativeness which has been tested, to representativeness where no test is possible.

If survey-makers do not take the necessary measures to secure the above conditions, they are not investigating by sample, and the only legitimate alternative is to make a complete enumeration either of the whole city, or, if resources are insufficient, of some continuous section of the city. No other procedure will yield results that can be expected to be representative.

In some surveys, unfortunately, the principles of sampling have not been followed. That of Buffalo, being recently published, may be taken as an illustration. In Buffalo, nine considerable areas, varied in character and scattered in position, were selected, and the main figures of the report are based on their aggregate. Now, there is no suggestion that the nine have any common distinguishing property, which would make the aggregate more than the result of arithmetic. The authors have been careful to point out that the areas have distinctive racial characteristics; the occupational distribution also is very varied, and there is much variation in idleness rates by race and by occupation. Therefore, the areas may quite conceivably furnish examples of all races and occupations, but there is no reason to suppose that the aggregate of the nine corresponds at all with the whole city as regards *distribution* either by occupation or by race. Statement is made on page 6 of the report that "it does not purport to show employment and unemploy-

ment for the whole city of Buffalo but merely for the persons enumerated." But one may fairly question the practical bearing of this statement. It is questionable whether in a report intended for a public not composed of statistical experts, such a single short sentence, however explicit, will exonerate the authors from responsibility, if in the ensuing publicity their 9.9 per cent idle is broadcast as the rate of idleness in the city of Buffalo; and such an outcome is certainly to be expected.

One cannot but feel that procedure of this kind shows insufficient consideration of the proper objective in such an investigation. If a study is not meant to be taken as representative of a whole city, but is an aggregate of districts with no evident relationship, what object does it serve? Examination of rates for districts will give some indication of variation within the city, but the indication is far from adequate, and has never been professed as the object of a study, nor does it seem a sufficient main objective. The Buffalo survey has indeed served a very useful purpose by proving that great variation occurs within a city. Male idleness in the Buffalo districts ranged between 6.9 and 16.2 per cent, and female idleness between 1.4 and 11.0 per cent. Evidently, when there is such variation among nine districts comprising a small portion of the city, the rate of idleness for the city as a whole remains practically as conjectural as before the investigation. But while this survey has great illustrative value, no such value would attach to subsequent investigations by the same procedure; further examples would be superfluous. The separate figures for these areas illustrate splendidly the futility of attempting to get representative figures by such selection. That futility has long since been demonstrated theoretically; here is spectacular practical demonstration for those who are impervious to abstract reasoning.

Avoidance of unsound procedure is particularly vital in the matter of sampling, because the justification for investigation by sample is not likely ever to be completely understood by the general public. In consequence, unsound results caused by substitutes for sampling are likely to be imputed to the method of investigation by sample instead of to its disregard, and may become disastrous obstacles to the winning of public confidence.

CONCLUSION

This article is emphatically not intended to suggest that all the points which it covers must be investigated and distinguished in all employment-unemployment surveys. Obviously, the alternatives will often be a less comprehensive survey or none at all, and a much less com-

prehensive survey may have great value. The vital necessity is that in planning a more limited survey all these points shall be considered, and should also be mentioned in the report so that the exact significance of the results be appreciable by users of the report. Selection of questions must be made with very great care so as to make the categories retained as useful as possible, and to present a logical whole. Sometimes the merging of two categories may make a result lack meaning; sometimes the omission of one question may remove the significance of another, which may therefore be as well omitted.

The essentials of an employment-unemployment survey may be recapitulated as follows: We need adequate employment classification of the enumerated persons, with particular emphasis on the necessity of keeping entirely distinct, if enumerated, any persons who have not been or no longer are usually employed for wages or salary. We need idleness rates based always on the proper totals, namely on the corresponding group of persons usually employed. We need either rigorous sampling preferably by households and never by units larger than blocks, or else complete enumeration of a continuous area. Any selection that is not rigorous sampling is not sampling at all, but is an aggregate which cannot be expected to be representative, and could not be proven representative even if by remote chance it should be so. In any case where observance of these conditions means non-comparability of the new with the old, we must remember the inevitable implication that the old must have lacked significance in one way or another, since the conditions are necessary for significance. The one defensible course is courageously to cut our losses. Let us by all means have more unemployment surveys, the more the better, but let us make them better as well as more.

PRESENT CONDITION AND FUTURE PROGRESS OF STATISTICS[1]

By Corrado Gini, *University of Rome*

The sciences, in their evolution, do not depart from that rhythmical movement which seems to dominate all the manifestations of cosmic forces, inasmuch as they alternate their activity between the collection of facts, on the one hand, and theoretic constructions, on the other. Thus, in the biological sciences, the period of the great theoretical constructions which take their name from Lamarck and Darwin was followed by an intense period of observations and experiments from which, in recent years, a new fervor in scientific systematizations has spread. Similarly, in physical and chemical sciences, after the formulation of the atomic theory, theoretic speculation had stood aside for experimentation, which has constituted the material for the recent construction of new systems.

Statistics also is naturally subjected to this rule that is common to all the sciences, but it presents special characteristics by which the rhythm, indeed, is intensified. As a matter of fact, the absence or inexactitude of the data collected is, in it, a stimulus to the progress of important branches of the theory.

Statistical methodology may indeed be divided into five parts, according as it treats: (1) of the collection of data; (2) of their tabulation and classification; (3) of the measure of the various aspects of the mass phenomena; (4) of the integration, correction and comparison of the data and measures obtained; (5) of the applications of the inductive and deductive methods to the data furnished by the above mentioned operations.

Now, it is clear that the processes involved in the development of the first and second parts, on the one hand, and those of the fourth and fifth parts, on the other hand, are, in a sense, complementary. Indeed, the less detailed and precise the data at one's disposal, the greater is the importance of the statistical induction which enters into the fifth part. And, above all, the more frequent and more important does it become to have recourse to methods of correction and integration of the data and to expedients necessary to their comparison.

All conjectural statistics has, for example, a *raison d'être*, in so far as the data directly observed are lacking. Its field, which once embraced

[1] Lecture given at the University of Toronto, March 25, 1930. Dr. Corrado Gini is Professor of Statistics at the Royal University of Rome and, at the same time, President of the Central Statistical Institute of the Kingdom of Italy.

almost the whole domain of statistics and contained in particular the calculation of population, has gone on restricting itself, according as the data of censuses have become more extended. There is, however, often occasion to apply it in other fields. This is especially the case in the calculation of the wealth and income of nations, where it has assumed the highest development because of the lack, in almost all States, of directly collected data.

Another chapter of statistical methodology, dependent on the limitation of collected data, is that of the dissection of curves, by which we wish to specify the part that two or more component curves have in determining the resultant curve. Such an elaboration might be avoided, if it were possible to indicate distinctly for each individual case the characteristics upon which the component curves are differentiated.

A great many of the researches which aim at establishing whether the differences among the data collected have a random character, have a justification only in so far as the observations do not extend to the universe; and this is also the source of all that chapter of statistical methodology which has for its object representative researches.

Tasks of interpolation, which object is to fill up the voids in series of data or to subdivide classes which are too comprehensive in character, would have, of course, no longer a *raison d'être*, if the classifications of the data were sufficiently detailed or the series complete.

Similarly, the application of the methods of elimination and of the process of multiple correlations would be avoided, if the data were sufficiently numerous and classified according to the presence of all those characters or phenomena whose effects we wish to examine separately.

Various other methods, more or less ingenious and more or less well founded, have been devised to remedy imperfections of data, and have a purpose only in so far as these imperfections exist. Such, for instance, is the method which presumes that it is possible to construct life tables without any knowledge of the number of those exposed to risk in each age class.

We certainly do not have to expect that, with the advancement made in compiling and classifying data, the other parts of statistical methodology are destined to become unimportant. Various reasons justify this position.

(a) First of all, presumably, there will always be in the world less advanced States in which the data collected and classified will contain gaps or be imperfect. Indeed, it is necessary to bear in mind that, above all, it is to remedy the gaps or imperfections which they present in States statistically better organized that there have been devised the methodological processes to which we have referred; and a long period

of time will have to pass before the States which do not yet possess statistics, or only possess them in an embryonic form, will achieve such an organization.

(b) On the other hand, methods of investigation based on incomplete data will always have a reason for their application as aids to the discovery of new facts and new connections, which have not as yet been subjected to systematic observation. This will be the case every time that the importance of a new field of research becomes appreciated.

(c) Again, it is to be noted that it will still be possible for certain statistical expedients to have, although in a more limited measure, an occasion for application, however extensive and exact the collection and however minute the classification of the data may be.

Such is the vast category of the processes which aim at measuring the influence of chance. They are processes which may certainly be rendered superfluous by the amplitude of the observations, when their object is to establish the presumable differences between the complete observations, which might have been made, and the incomplete ones, which have actually been made. They cannot, however, be dispensed with, when they aim at measuring the probable differences between the results of complete observations and the results, which would have been obtained from hypothetical observations, that might have continued indefinitely in identical circumstances of time and place.

Similarly, the methods of interpolation will always be rightfully applied, in so far as they aim at deducing, from the complete results of observation, the laws which govern the phenomena, and at representing with synthetic formulas, for the convenience of study, huge masses of observed facts. Similarly, one will be obliged to have recourse to the methods of elimination every time one wishes to get rid of the influence of factors which intervene in a restricted number of observed cases.

(d) It is also to be observed that the progress of that part of statistical methodology, which has to do with the measure of the various aspects of mass phenomena and which still has regions open to development, carries along with it the advance of that part which treats of the comparison of the measures obtained and also the advance of that part which has to do with statistical induction. As a matter of fact, it offers, to these, new material for application. Thus the introduction, into statistical methodology, of the concepts and measures of the transvariation and of the dissimilarity of distributions, brings with it the examination of the probable errors by which, owing to chance, the respective indexes are affected.

It is certain, however, in any case, that the parts of statistics which have to do with the collection and classification and tabulation of data,

are a long way behind those which treat of the measure of mass phenomena, of the integration, correction and comparison of the results and of methods of reasoning. This is because the development of these latter parts has been brought about by scientists, whereas that of the former parts has up to now depended to a greater extent on administrations. As a matter of fact, the scientists in most States have heretofore either taken no interest in these matters, or, if they did intervene, it was only spasmodically, and usually in a merely consultative capacity.

The disadvantage resulting from this state of affairs is twofold. On the one hand, we have here—as I said—one of the reasons why the collection and classification of data are still in a backward state. On the other hand, an erroneous conception of statistics has thereby been facilitated. This conception is not uncommon with statisticians whose training has been emphasized on the mathematical side. From their viewpoint, statistics is regarded as a formal science, which is an end in itself, and not as a technique, whose devices must be judged solely in relation to their applicability to concrete facts and by the results which may be derived from them. The construction of ingenious methods, impossible of or at least difficult of application, is indeed to be discouraged by those who properly understand the function of statistics. Between two statistical methods which lead to the same result by unequally complicated ways, the simpler process is always to be preferred, even if the more complicated one gives those who apply it an opportunity of showing their exceptional technical abilities. In the same way, those instruments which are used for cultivation of the fields, for astronomical observations, or laboratory experiments, are valued not because of the complexity and ingeniousness of their construction, but by virtue of facility and fruitfulness of their applications; so must statistical processes be judged.

Not "Statistics as Mathematics," not "Statistics with Mathematics" should be, as I elsewhere observed,[1] the motto of those who understand the purposes of statistics, but "Statistics with the least mathematical means possible."

A better coördination of the compilation and classification of data with the more elaborated parts of statistical methodology will then be of advantage even to those who would promote the more refined statistical processes, for it would continually remind them of the ultimate purpose of the apparatus they are preparing. It is certain, however, that, above all, the stages of the collection and classification of data will derive the greatest profit from such a coördination.

[1] *Cf.* "The contributions of Italy to Modern Statistical Methods," *Journal of the Royal Statistical Society*, July, 1926.

To this end scientists should be employed to a greater extent in statistical bureaus, and with duties of direct responsibility rather than in a consultative capacity, for it is not so easy for one who has not previously borne such responsibility to give competent advice. All this is the more necessary, since the enlargement of statistical bureaus and the extension of their functions, coupled with the growth of the science of statistics, have rendered less and less common outstanding chiefs of statistical bureaus who combined administrative abilities with scientific abilities, and who so largely contributed in the past to the progress of statistics, in its theoretical and applied phases.

The difficulties that must be overcome are not slight. Statistical bureaus are not scientific bureaus, but are rather administrative bureaus. The necessity for exactness in the data and the requirements of their classification are felt therefore only in regard to the administrative uses which may be made of them. The personnel generally does not look upon them as constituting an instrument for a superior function—the registration of vital phenomena of the nation—the object of which is to give to the government the elements that are necessary for directing its affairs.

Furthermore, the importance of the prompt collection and publication of statistics is felt by each administration in regard to the particular end which it has in view, and, since very often for this particular end other official requirements may be of more importance, it results generally in the sacrifice of statistical observations, whenever it is found impossible to handle statistical and administrative affairs with the desired rapidity.

With the institution of central bureaus of statistics, which is now an accomplished fact in many States, and toward which there is a clear tendency in several other States, an improvement has already been brought about. As a matter of fact, a coördination amongst the compilations of the various administrations, which was frequently lacking, has thus been realized. The training of an expert personnel in the field of statistics, whose career can be developed and completed within the same central bureau, is thus facilitated. And furthermore it has resulted in the compilation and elaboration of data being frequently inspired by ends beyond those of any one department.

These advantages, however, relate exclusively to the function of the centralization and elaboration of data, whilst, with one sole exception, to which we shall refer later, the gleaning of data is entrusted to the personnel of other administrations. Central bureaus of statistics are indeed in a singular situation, inasmuch as, differing from all other central administrations, they have not their own circumferential

personnel, but make use at the periphery of the personnel of other administrations. The system has evidently an historical reason, inasmuch as the service of statistics arises for solely administrative purposes, as an appendix to the other functions of the State or of local bodies. The inconveniences of the system are evident. I have already mentioned some of them in connection with the feeling of the exactness of data and of the requirements of their classification and in connection with the rapidity of their collection, elaboration and publication. Another very serious inconvenience arises from the fact that the works, which the central bureaus of statistics must answer for, thus find themselves practically at the mercy of other administrations.

The system adopted in Russia by the Soviets aims at remedying these objections. In the Union of Soviet Republics, statistics has quite a special development. This must be placed in relation to the communist character of the State, which brings it about that statistical data assume a still greater importance than in States with a liberal or corporative régime, because it is to the State that the initiative and responsibility of the economic development of the nation belong. It is not therefore to be wondered at, that statistical services constitute there a ministerial department whose chief forms part of the Council of the Commissaries of the People, which is equivalent to our Council of Ministers.

But still more radical is the difference which comes from the fact that in Russia the Central Bureau of Statistics has at its disposal an army of specialized functionaries scattered all over the Republic, whose object it is to collect, at least in great part, those statistical data which are supplied in the other States by the employees of the circumferential bureaus of other administrations or of local bodies (in Italy, for example, communal secretaries, bureaus of the prefecture, provincial bureaus of economy, bureaus for raising taxes, registry offices, travelling chairs of agriculture, etc.).

Whilst the advantages of the system are evident, not less evident, however, are its disadvantages. These disadvantages arise, on the one hand, from the high cost of maintenance. In fact, statistical service costs in Russia 50 per cent more per head than it costs in other States which pay more for it, and over six times what it costs in Italy which is among the States which spend the least. On the other hand, they arise from the impossibility of finding, in sufficient numbers, a personnel conversant with all the branches of statistical registrations and with the working of the wheels of administration, so as to be aware of the deficiencies which statistical data may present by reason of administrative arrangement. This is a very important quality in order

to discover and correct the errors pertaining to them at the very source.

If the means at the disposal of Governments for statistical services increase and the diffusion of statistical knowledge becomes general, it is not impossible for the Russian system to have the future for itself. At present, however, although we recognize it as having certain advantages, it does not seem that one can advise its adoption.

But, even with the usual statistical organization, essential progress may be made, above all in certain directions.

First of all, as regards scrupulousness in the collection of data; if many scientists, who worry their brains in excogitating hair-splitting methods, had in view the deficiencies of the data to which they themselves apply or would have others apply the methods, they would be so overcome as to interrupt their labors.

Inexactitudes may be concerned with the very entity of the data, as happens, for example, in declarations of age which, as is known, tends to be given in round figures; they may be concerned with the time to which statistical data relate, as declarations of the date of birth, which in several States are kept back to the end of the year, in order to make children appear younger; and, finally, with the characteristics which relate to statistical unity, as happens with causes attributed to deaths or with professions registered in the census.

The seriousness of the trouble comes from the fact that, whilst the arguments bear obviously the mark of their worth or their falsity, so that, even if for a certain period they succeed in deceiving slack human intellects, they end up by being valued for what they are worth, in statistics, on the contrary, it is extremely difficult, except in particular cases, to separate good figures from false. Moreover, it seems that numerical data exercise over the human intellect a kind of fascination by which the very ones who make them, and should be more or less conscious of their defects, take them then for liquid gold, and claim, to use the old phrase, to give the exact weight of that material with the goldsmith's balance, which no one better than themselves should know can be weighed only by the steelyard of the charcoal dealer. Such, for example, is the case of many doctors who, whilst they are responsible for the inexactitudes which happen in the declarations of the causes of deaths, do not hesitate to entrust themselves to the slightest differences of statistics to find therein the proof of their hygienic or therapeutic theses.

Only by generally diffusing and implanting in the functionaries and non-functionaries that statistical conscience, which springs from the consciousness of national ends served by statistics, can we efficaciously

eliminate this objection. But the same may be said of public conscience that has been said of divine providence, that it likes to be helped by the initiative of those who have the responsibility and interest of the operation. In our case, the conscience of the personnel may be roused by means of special courses of instruction under the direction of central bureaus of statistics (like those which have just been started in Italy). It may be also assisted by means of subsidies for statistical compilations paid by central statistical bureaus to the personnel of other administrations and likely to remind it of the general utility which data have, apart from that which they assume for each single administration. And the consciousness of functionaries and non-functionaries may be moreover facilitated in its manifestations by means of a suitable drawing up of questions which should contain a control in themselves, as, for instance, when age is asked for, it should be accompanied by the question of the year of birth. It may be finally superintended by controls aiming to fill up the gaps and eliminate errors in the data furnished; controls made either at the periphery, as in Italy in the bureaus of the Prefecture, or at the center, as is the case in every well organized central bureau of statistics. It is obvious that the efficacy of these controls depends on the conscientiousness of those who are placed over them; but, of course, it is much easier to find a maniple of scrupulous controllers than an army of scrupulous compilers.

Another field, in which a decisive advance is to be expected, is that of the classification of data which, whilst it must keep in view the administrative ends which statistics serves in its drawing up and in its particulars, must, moreover, correspond both to the ends of science and to the necessities of comparison with other national and international statistical data.

It is, of course, difficult for a country to abandon the criteria of classifications which it considers most suitable and in a certain way adherent to its administrative or political organization, in order to obtain comparison with the data of other States. In some fields, however (for instance criminality or tax pressure) in which the estimate of the phenomena assumes a particular importance in comparison with other nations, such a necessity becomes evident; nor are there lacking, moreover, international organs which go deeply into the question of such comparability.

However great the importance of this international comparability may be, the importance of the connection of the classifications of statistical data with national scientific exigencies is still certainly greater. Nor should it cause surprise that in this field one is still far from the desired perfection, in spite of the advice which is given in all

States, by means of superior councils or special commissions, by the best promoters of academic statistics. Very often, indeed, they, either through ignorance of administrative organizations, or through an omitted consideration of their financial requirements, or through an innate deficiency of a practical mind, or owing to that kind of professional deformation, of which we have spoken, which leads them to look upon statistical methodology as a purely formal science, they, I say, make proposals so removed from all practical execution, that even the germ of good enclosed within them is rendered unproductive. Nor are there lacking at times statisticians who are so attached to methods to remedy the imperfections of statistics that it is with evident disappointment they witness those improvements in collecting and classifying data, which would render such methods superfluous.

Another source of advance in statistics can come from a more thorough scientific elaboration of collected and classified data on the part of bureaus of statistics. Having tested the building material and being conversant with it, these bureaus are often in a position to use it, mindful of the different solidity of its various elements, better than eminent scientists who have only the superficial knowledge of those who see it from the outside. This naturally presupposes that bureaus have a personnel capable of presenting in scientific reports the material that has been worked out. Certainly one of the most common defects of central bureaus of statistics consists in the lack of such a personnel.

In the organization of the Central Institute of Statistics in the Kingdom of Italy I have tried to remedy it by the institution of a special department of studies and of a special department of mathematical researches. The presence of a scientific personnel in subordinate posts of command is even more desirable perhaps than at the head of the bureau itself. To have at the head of the Central Bureau of Statistics a scientist is a condition to be greatly desired in principle, but which, in practice, is difficult to realize, and, in fact, is realized in the statistical bureaus of the world only exceptionally. As a matter of fact, it is difficult to meet with the happy combination of a scientist who has the administrative capacity indispensable for the management of a central bureau of statistics, and those endowments, equally important, of energy which, with the present arrangement, are necessary to be able to make a bureau work efficaciously whose circumferential organs depend on other administrations. And perhaps it is still more difficult to find, in addition to such qualities, that spirit of abnegation which permits him to neglect, if not to abandon altogether, for a long number of years, with a sacrifice rarely adequately appreciated, his favorite occupations for duties, perhaps more important from an immediate point of

view, but certainly not of greater definite importance, and, anyway, not corresponding to his subjective preferences.

The advantages of a greater scrupulousness in the collection and control of data, of a more detailed and scientific classification and of a more thorough elaboration, seem however to be badly reconciled with another requirement of statistics, too often neglected, i.e., rapidity in publication. On this point also a substantial progress on the part of many central bureaus of statistics is still a desiderata.

For it is natural that, if there is a delay in publishing statistics, they almost completely lose their interest, both for the government men and for the business men, who precisely represent the two categories of consumers of statistics which can efficaciously contribute to finance or get financed the central bureaus of statistics (for one can count very little, in this respect, upon the third great category, namely scientists). There is no other solution in this respect than that of organizing statistical collections and elaborations so that a first provisory result, limited to those substantial indications which may interest business and government men, should be published with all haste, whilst definite data with scientifically desirable classifications, elaborations and illustrations, will be published only later. This system has already been adopted in Italy for the publication of data on births, marriages and deaths and, beginning with this year, it has been also adopted for those relating to migratory movements. Summary data are collected from month to month so that within the following month one is in a position to know the numerical variations of the population of Italy and of each single province. Data giving a more detailed classification are then published five or six months after the closing of the year; finally, definite reports drawn up on schemata prepared by specialists, are completed at least a year later. This same principle will be adopted as regards the general census of the population, as it is planned that essential data will be given out to the public within a year from the date of registration, whilst the more detailed elaboration and reports will appear later, under the care of a special department that will become permanent.

Statistics has a rather humble past, but a very brilliant future; and, as all those things, which have a poor past and a promising future, it has many difficulties to overcome.

NOTES

ILLUSTRATION OF THE APPLICATION OF A COEFFICIENT MEASURING THE CORRELATION BETWEEN A VARIABLE AND THE DEVIATION OF AN ASSOCIATED BUT NOT DEPENDENT VARIABLE FROM ITS PROBABLE VALUE

By J. Arthur Harris[1] and Borghild Gunstad, *University of Minnesota*

In this Journal for 1922, one of us gave the formula for a coefficient measuring the relationship between a variable and the deviation of an associated but not dependent variable from its probable value,[2] thus extending earlier work[3] on the same general subject. At that time no illustration of the applicability of the method in the field of social or economic statistics was available, and the formula has stood without an example of numerical results.

In an investigation of fecundity in foreign-born and American-born inhabitants of the City of New York it was found that the proportion of pregnancies brought to term was higher in foreign-born than in American-born mothers. In the analysis of the result it therefore became necessary to determine whether the differences in the capacities of the women of the several nationalities for bringing their pregnancies to term might be definitely related to the ages of the mothers at the time at which the records of numbers of pregnancies and births were secured—that is, at the time of the birth of the last recorded child.

This problem may be solved by determining the correlation between the age of the mother and the deviation of the number of births from their probable value by the formula

$$r_{wz} = \frac{S\{w[\Sigma(y) - p\Sigma(x)]\}/N}{\sigma_w \sigma_z}$$

where w = age of parents, y = number of pregnancies brought to full term and x = number of pregnancies incurred, σ_w is the standard deviation of age and σ_z is the standard deviation of the deviation of the num-

[1] Published posthumously.

[2] J. Arthur Harris, "On the Correlation Between a Variable and the Deviation of an Associated but Not Dependent Variable from Its Probable Value," this Journal, 18: 393–394, September, 1922.

[3] J. Arthur Harris, "The Correlation Between a Variable and the Deviation of a Dependent Variable from Its Probable Value," *Biometrika*, 6: 438–443, 1909.

ber of births from their probable value as defined elsewhere,[1] Σ denotes summation for the array of x or y associated with any age class and S denotes summation for all classes of w, or for the N individuals of the sample.

The data analyzed represent the histories of 11,011 women confined at the Sloane Hospital for Women, New York.[2] Information as to the birthplace of both parents was secured by culling from 44,493 official birth certificates for the period September, 1890, to September, 1921, all foreign-born mothers of certain nationalities whose husbands were also born in the same foreign country, and a roughly parallel series of records of American-born mothers with American-born husbands.

Since a woman whose current pregnancy has been terminated by a miscarriage does not visit the hospital for confinement, the data as obtained at a hospital for the last recorded pregnancy and confinement do not furnish an accurate record of the percentage of miscarriages at that pregnancy. The number of earlier pregnancies and miscarriages are, however, recorded at the time of her entrance to the hospital. We may regard these records as much more nearly representative of the general population.

The correlations in Table I are given for the original series, including the last (here always fruitful) pregnancy, and for the reduced series, excluding the last pregnancy.

The correlations between age and number of pregnancies and between age and number of births are, in the larger series, of the order .45 to .60. In general the coefficients are slightly higher in the original than in the selected series. The differences are, however, very small. The medium values of these coefficients attest the large influence of social and economic conditions on human reproduction. If it were a purely biological phenomenon the correlation between age and number of pregnancies and between age and number of births would be expected closely to approach 1.00.

The correlations between age and the deviations of the number of births from their probable value are in all cases small and are in part positive and in part negative in sign. Thus, in the original series 8 of the 10 constants are negative, whereas in the reduced series 8 of the 10 coefficients are positive. Possibly further research may indicate a trustworthy relationship between these two variables. As far as the data go they show that there is no large relationship between the age

[1] J. Arthur Harris, "Further Illustrations of the Applicability of a Coefficient Measuring the Correlation Between a Variable and the Deviation of a Dependent Variable from Its Probable Value," *Genetics*, 3: 328–352, 1918.

[2] For permission to use these records we are indebted to the late W. E. Studdeford, then Superintendent of the Hospital.

of the mother as here defined and the proportion of pregnancies brought to term.[1]

While the present constants are given primarily as an illustration of a statistical method, the results derived from the present data are ample to indicate that such relatively slight differences as may be found between the mean ages of the mothers of different nationalities will not introduce any serious error in the comparison of the ratios of births to pregnancies in these nationalities.

TABLE I

CORRELATIONS BETWEEN AGE OF MOTHER AND NUMBER OF PREGNANCIES, r_{us}; AGE OF MOTHER AND NUMBER OF BIRTHS, r_{uy}; AND BETWEEN AGE OF MOTHER AND DEVIATION OF NUMBER OF BIRTHS FROM THEIR PROBABLE VALUE, r_{us}

Nationality	Original Series				Reduced Series			
	Number of mothers	r_{us}	r_{uy}	r_{us}	Number of mothers	r_{us}	r_{uy}	r_{us}
Austrian	253	.549	.555	−.123	143	.477	.486	−.023
English	328	.461	.475	+.020	201	.490	.514	+.155
German	845	.565	.553	−.095	514	.483	.478	+.078
Greek	88	.330	.343	+.008	47	.110	.141	+.114
Irish	2,208	.572	.563	−.130	1,336	.556	.530	−.005
Italian	249	.503	.504	−.019	182	.538	.539	+.111
Russian	625	.701	.684	−.091	338	.672	.649	+.081
Total Foreign	4,596	.565	.553	−.097	2,761	.578	.562	+.041
American	6,415	.529	.499	−.148	3,629	.498	.459	+.051
American and Foreign	11,011	.544	.533	−.084	6,390	.529	.521	+.112

[1] The correlation between age and the deviation of the number of births from their probable value as here determined cannot be considered a satisfactory measure of the relationship between age and fecundity, or the ratio of the number of offspring brought to term to the number which might have been brought to term. This is true because our measure of age is necessarily the age of the mother at the birth of the last child. Paired with this variable we have merely the total number of pregnancies and births occurring during an antecedent period of unknown duration. For example, it is quite possible that of two mothers each 35 years of age and each having been pregnant three times and each having had one miscarriage, one may have had her first two pregnancies and her one miscarriage between the ages of 18 and 23 years of age while the other had her first two pregnancies and her one miscarriage between the ages of 31 and 34 years. These two biologically quite different categories cannot be distinguished on the basis of data reporting only the age of the mother at the birth of a given child and the number of preceding pregnancies and miscarriages. The problem is an important one, physiologically and socially. We much need data which will show the ages of the mothers at all incidences of pregnancy together with the fate of each pregnancy.

WHOLESALE COMMODITY PRICE INDEXES—WEEKLY, MONTHLY, QUARTERLY, AND ANNUAL

SEVENTH REPORT

By Irving Fisher

A description of my revised wholesale commodity price indexes calculated on the base 1926=100 is given in this Journal for December, 1928. As explained in the article referred to, the base year (1926=100) and the quantity weights (average quantities marketed in 1923 and 1925) for the new indexes were made to conform to those of the United States Bureau of Labor Statistics indexes. The weekly index is therefore designed to be an indicator of the Bureau's monthly index.

Since the last report was made, no changes of importance in the methods of collecting the prices or calculating the indexes have been made.

Group indexes have been computed each week since the revised index was begun on January 6, 1928. The groups composing my index are identical with those of the United States Bureau of Labor Statistics index, except that the group, house furnishing goods, has been omitted because no satisfactory price quotations were available. The total value of these goods is very small being only 1.2 per cent of the value of all commodities in the United States Bureau of Labor Statistics index for 1926.

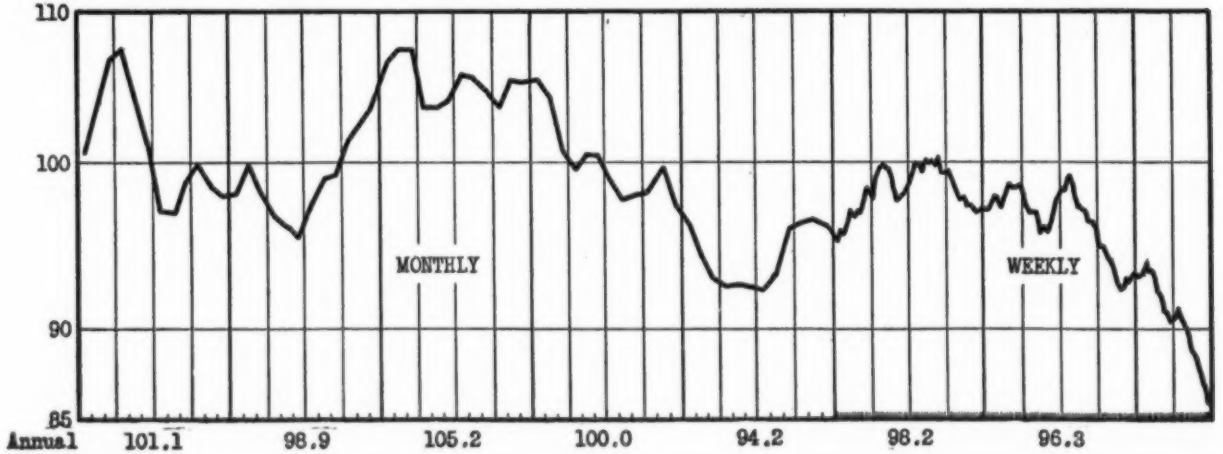
The number of commodities included is 120. These are apportioned among the several groups as indicated in the tables below. The price movements of these articles tally very closely with the price movements of the 550 commodities included in the indexes of the United States Bureau of Labor Statistics. Each month the all commodity and the group indexes are adjusted to conform to a three months' moving average of the Bureau's indexes.

The accompanying tables carry the revised indexes forward from September, 1928, and back to 1923, the year in which these indexes were established. Group indexes were first constructed beginning with 1928 and no weekly indexes have been computed on the base 1926=100 back of that year.

TABLE I

WHOLESALE PRICE INDEXES ALL COMMODITIES (1926=100)

Month	1923	1924	1925	1926	1927	1928	1929	1930
Jan.	100.6	98.0	106.7	105.4	96.0	95.5	97.4	93.3
Feb.	103.2	99.9	107.5	104.2	94.4	96.7	97.7	92.7
Mar.	106.7	98.0	107.3	101.0	92.9	97.6	98.3	90.8
1st qtr.	103.5	98.6	107.2	103.5	94.4	96.6	97.8	92.3
Apr.	107.6	96.6	103.5	99.6	92.5	99.1	97.1	90.6
May	104.5	96.0	103.5	100.5	92.5	99.1	95.9	88.6
June	101.2	95.4	104.0	100.5	92.4	98.0	96.9	86.4
2nd qtr.	104.4	96.0	103.7	100.2	92.5	98.7	96.6	
July	97.0	97.3	105.8	98.8	92.2	99.6	98.5	
Aug.	96.9	98.9	105.6	97.5	93.2	99.8	97.3	
Sept.	98.8	99.2	104.6	97.9	95.8	99.7	96.0	
3rd qtr.	97.6	98.5	105.3	98.1	93.7	99.7	97.3	
Oct.	100.0	101.4	103.6	98.0	96.2	98.8	94.4	
Nov.	98.4	102.6	105.2	99.7	96.4	97.5	92.7	
Dec.	97.8	103.5	105.2	97.3	96.1	97.0	92.8	
4th qtr.	98.7	102.5	104.7	98.3	96.2	97.8	93.3	



Annual	101.1	98.9	105.2	100.0	94.2	98.2	96.3

TABLE II

WHOLESALE COMMODITY PRICE INDEXES (1926=100)

1928

WEEKLY

Week ending	All products (120)	Agricultural products (30)	Non-agricultural products (90)	Raw materials (34)	Semi-mfd. goods (22)	Finished goods (64)	Farm Products (15)	Food-stuffs (19)	Hides and leather (8)	Tex-tiles (16)	Fuel and light (10)	Metal products (18)	Build-ing material (20)	Chem-icals (14)	Mis-cella-neous (7)
Jan. 6	95.5	102.4	95.5	102.2	97.3	95.8	107.7	100.6	128.8	98.4	84.6	96.9	89.0	97.2	87.2
13	95.1	101.6	95.9	101.3	97.6	95.8	107.1	100.9	135.8	97.6	84.6	96.9	89.1	96.7	86.8
20	95.7	101.0	96.5	101.7	97.8	95.7	106.4	100.8	136.2	97.6	85.8	97.1	89.2	96.6	87.6
27	95.5	100.4	95.4	99.9	97.2	95.9	105.7	101.2	135.1	97.0	83.1	97.2	89.7	95.1	85.9
Feb. 3	96.0	99.7	95.6	100.1	97.8	94.6	104.7	100.1	136.1	96.5	83.2	97.7	90.1	94.6	85.6
10	97.0	100.0	95.6	100.5	97.6	94.1	105.7	100.0	134.7	96.2	83.1	97.9	89.8	94.1	85.5
17	96.8	99.6	95.4	99.9	97.8	93.9	104.9	99.9	132.0	95.8	83.3	98.0	90.4	94.3	85.1
24	96.6	98.7	94.8	99.4	97.5	92.8	104.0	98.7	129.3	95.8	82.7	97.9	90.5	93.7	84.3
Mar. 2	96.9	100.0	94.5	99.8	97.6	93.3	105.1	99.5	127.9	95.9	82.0	97.8	90.6	93.6	84.5
9	96.8	99.0	94.1	99.3	97.5	92.3	102.7	98.6	126.5	95.9	81.5	98.2	90.2	93.0	83.9
16	97.3	99.5	93.8	99.7	97.5	92.2	104.1	98.8	125.5	95.5	81.0	98.2	90.4	93.2	83.8
23	98.2	99.9	93.7	99.6	97.7	92.8	104.3	99.4	123.3	95.9	80.8	98.4	90.5	93.2	84.4
30	98.1	99.9	93.7	100.2	97.6	92.6	105.4	98.6	123.6	95.2	80.6	98.5	90.5	93.6	85.9
Apr. 6	97.6	99.2	93.2	99.6	97.3	92.0	105.0	97.5	123.1	95.2	79.8	98.7	90.4	93.8	84.8
13	98.9	101.4	93.8	101.8	97.3	93.0	107.6	98.5	128.9	95.1	80.7	98.3	90.4	93.3	84.4
20	99.4	102.9	94.3	102.8	97.3	93.6	109.5	99.2	129.1	95.8	80.7	98.7	90.5	93.5	84.5
27	99.7	104.7	94.0	103.6	97.1	92.5	111.8	100.0	128.8	96.1	80.3	98.3	90.4	93.5	84.3
May 4	99.8	104.9	94.0	103.7	97.3	94.6	112.2	99.4	128.6	96.8	80.3	98.2	90.6	93.9	84.8
11	99.6	105.0	93.9	103.7	97.3	94.4	112.8	99.2	128.2	96.4	80.3	98.0	90.5	94.3	85.5
18	99.6	105.5	93.9	105.0	97.2	95.5	111.7	99.9	128.0	96.3	80.3	98.3	90.6	93.9	85.8
25	98.8	104.5	93.8	102.0	97.3	95.5	110.0	99.8	126.0	96.4	80.4	98.3	90.8	94.1	85.7
June 1	98.0	103.6	93.7	100.4	97.6	95.8	107.6	98.8	123.1	95.5	80.6	98.4	91.0	94.6	86.7
8	97.6	103.6	93.9	99.9	98.0	95.8	107.2	98.5	122.5	96.0	80.7	98.1	91.2	94.4	87.5
15	97.8	102.9	95.5	101.1	97.8	95.8	105.5	98.0	122.7	98.1	84.6	98.0	91.5	94.3	86.5
22	97.9	103.6	95.6	100.7	98.1	96.6	105.0	97.8	122.7	98.2	84.7	97.7	91.8	94.4	86.9
29	98.4	105.1	95.8	101.1	98.2	97.2	106.2	98.3	122.7	98.8	84.8	97.8	92.1	94.3	86.5
July 6	98.8	106.2	96.1	102.3	98.6	96.9	107.9	97.8	123.5	99.1	84.9	97.9	92.4	94.5	86.9
13	99.3	107.4	96.4	102.3	98.9	98.1	108.0	99.1	126.2	96.6	85.9	97.4	92.7	94.6	87.6
20	99.9	109.0	96.5	102.2	98.8	99.9	109.8	101.8	126.5	96.5	85.8	97.4	93.0	94.9	87.0
27	99.9	108.5	97.0	101.7	99.5	100.1	109.0	101.8	126.4	97.6	86.6	97.5	93.2	94.9	86.5
Aug. 3	99.7	108.3	96.8	100.7	99.3	100.5	107.5	102.5	125.2	97.4	86.5	97.6	93.4	95.0	85.9
10	99.4	108.3	96.3	100.3	98.6	100.4	107.6	103.0	122.5	96.9	85.7	97.7	93.3	94.8	85.3
17	100.1	110.4	96.2	101.3	98.5	100.8	110.4	103.7	121.6	96.7	85.5	97.7	93.4	95.6	84.8
24	99.9	110.8	96.1	101.2	98.3	100.8	100.5	104.1	121.5	96.2	85.3	97.9	93.5	95.4	84.4
31	100.0	111.5	96.1	101.5	98.1	100.9	111.1	104.5	121.7	96.3	85.3	97.5	93.9	95.3	84.4

													97.8	93.9	95.3	84.4

Sept. 7	99.7	110.4	95.9	102.3	97.2	99.7	110.1	103.1	119.2	96.1	85.2	97.5	94.4	95.4	83.7
14	100.3	111.0	95.8	103.0	97.1	100.4	112.9	104.7	119.3	95.6	85.0	97.7	94.4	95.2	83.4
21	99.8	108.6	96.2	102.2	97.1	99.9	110.5	104.1	119.4	96.1	85.7	98.4	94.6	95.4	82.4
28	99.3	107.1	96.2	101.6	97.2	99.4	109.2	103.6	117.0	96.1	85.4	99.1	94.7	94.5	82.9
Oct. 5	99.2	106.2	96.2	100.9	96.8	99.7	107.9	104.2	114.1	97.4	85.2	100.2	94.8	94.2	82.1
12	99.4	106.0	96.1	100.5	96.6	100.1	107.7	104.8	111.3	97.1	85.7	100.6	94.8	93.8	81.6
19	98.9	104.3	96.0	98.8	96.6	100.3	105.9	105.5	110.4	97.0	85.1	100.5	94.8	94.0	81.6
26	98.4	101.9	96.4	98.2	96.6	99.2	104.4	103.7	109.9	98.0	85.3	101.2	94.8	94.1	81.5
Nov. 2	97.9	100.2	96.6	98.0	96.6	98.1	103.5	102.4	109.8	97.8	85.3	101.8	95.1	94.1	81.3
9	97.6	98.8	96.8	98.1	96.5	96.7	103.7	100.7	109.6	97.8	85.3	101.8	95.6	93.4	80.9
16	97.7	98.8	96.9	98.0	96.4	97.1	103.1	101.4	108.8	97.9	85.3	102.0	95.5	92.4	80.5
23	97.3	97.3	97.2	97.6	96.5	96.4	101.8	100.6	109.4	98.9	85.3	102.3	95.7	92.4	80.4
30	97.3	97.2	97.1	98.0	96.3	96.2	102.4	100.3	110.6	98.8	85.3	102.7	95.5	92.9	79.0
Dec. 7	97.1	97.4	97.1	97.4	96.7	96.3	102.0	100.4	109.0	97.5	85.3	102.7	96.7	93.0	80.5
14	96.8	96.6	97.6	96.8	97.7	95.5	101.5	99.5	109.5	97.3	85.4	104.4	97.3	93.3	80.5
21	96.9	96.5	97.7	96.6	98.3	95.5	100.8	99.3	109.4	97.1	85.3	104.2	96.8	93.4	83.4
28	97.2	96.8	98.0	97.7	98.5	94.9	101.1	98.5	110.1	99.2	85.1	104.5	96.8	94.1	83.4
MONTHLY															
Jan.	95.5	101.0	95.8	100.9	97.6	95.6	106.3	100.8	134.9	97.4	84.3	97.1	89.4	96.0	86.6
Feb.	96.7	99.6	95.2	100.0	97.7	93.7	104.9	99.6	132.0	96.0	82.9	97.9	90.2	94.0	85.0
Mar.	97.6	99.6	93.8	99.7	97.6	92.5	104.2	98.8	124.8	95.6	81.0	98.3	90.4	93.3	84.5
Apr.	99.1	102.6	93.9	102.3	97.3	93.5	109.2	99.0	127.8	95.7	80.4	98.5	90.5	93.6	84.5
May	99.1	104.7	93.9	103.0	97.3	95.2	110.8	99.4	126.7	96.2	80.4	98.2	90.7	94.2	85.8
June	98.0	104.0	95.3	100.8	98.1	96.4	106.1	98.1	122.7	97.9	83.8	97.9	91.7	94.4	86.8
July	99.6	108.0	96.6	101.9	99.0	99.2	108.6	100.8	125.8	97.3	86.0	97.5	92.9	94.8	86.9
Aug.	99.8	110.0	96.3	101.0	98.5	100.7	109.6	103.7	122.2	96.6	85.6	97.7	93.5	95.2	84.9
Sept.	99.7	109.2	96.0	102.2	97.2	99.8	110.6	103.9	118.6	96.0	85.3	98.2	94.5	95.1	83.0
Oct.	98.8	104.0	96.2	99.3	96.6	99.6	106.0	104.3	111.1	97.4	85.3	100.8	94.9	94.0	81.7
Nov.	97.5	98.2	97.0	97.9	96.5	96.8	102.8	100.9	109.6	98.3	85.3	102.2	95.5	92.9	80.3
Dec.	97.0	96.8	97.6	97.1	97.8	95.8	101.3	99.3	109.4	97.8	85.3	104.0	96.9	93.5	82.0
QUARTERLY															
1st quarter	96.6	100.0	94.9	100.2	96.9	93.9	105.1	99.7	130.6	96.3	82.7	97.8	90.0	94.4	85.4
2nd quarter	98.7	103.8	94.3	102.0	97.6	95.0	108.7	98.8	125.8	96.6	81.5	98.2	91.0	94.0	85.7
3rd quarter	99.7	109.1	96.3	101.7	98.2	99.9	109.6	102.8	122.2	96.6	85.6	97.8	93.7	95.0	84.9
4th quarter	97.8	99.7	96.9	98.1	97.0	97.4	103.4	101.5	110.1	97.8	85.3	102.3	95.8	93.5	81.3
ANNUAL															
1928	98.2	103.1	95.6	100.5	97.4	96.6	106.7	100.7	122.1	96.9	83.8	99.0	92.6	94.2	84.3

TABLE III

WHOLESALE COMMODITY PRICE INDEXES (1926=100)

1929

WEEKLY

Week ending	All products (120)	Agricultural products (30)	Non-agricultural products (90)	Raw materials (34)	Semi-mfd. goods (22)	Finished goods (64)	Farm Products (15)	Foodstuffs (19)	Hides and leather (8)	Textiles (16)	Fuel and light (10)	Metal products (18)	Building material (20)	Chemicals (14)	Miscellaneous (7)
Jan. 4	97.0	96.2	97.9	97.6	98.6	94.2	100.9	97.7	108.7	99.0	85.1	105.1	97.0	94.3	83.5
11	97.1	96.1	98.3	97.3	100.2	93.9	101.0	97.1	107.2	98.5	84.9	105.1	97.0	94.1	89.2
18	97.4	97.2	98.0	96.5	100.4	95.5	101.8	99.2	105.8	98.2	84.9	104.9	96.9	93.9	89.5
25	97.8	99.7	97.1	97.6	99.7	96.2	105.0	100.0	102.6	96.7	84.6	105.1	97.0	94.5	88.6
Feb. 1	97.7	100.3	96.4	96.5	99.5	97.5	104.4	101.4	100.8	96.1	83.5	104.9	97.0	94.9	90.0
8	97.1	99.5	96.1	95.5	99.4	97.3	102.4	101.0	99.2	95.2	83.3	106.0	97.2	95.0	89.9
15	97.6	99.9	96.5	97.6	98.9	96.8	103.3	99.8	94.1	94.8	87.0	106.0	96.7	95.4	89.8
22	98.0	100.9	96.4	98.0	98.7	97.7	105.1	100.9	93.9	94.7	87.1	105.8	96.6	95.6	89.5
Mar. 1	98.4	101.2	96.6	98.4	100.0	98.1	105.3	101.0	95.9	94.6	87.0	107.2	96.6	95.7	89.2
8	98.2	101.6	95.8	98.8	97.6	98.1	106.1	100.8	90.2	95.0	86.9	107.3	96.1	95.6	86.5
15	98.2	101.3	95.7	98.8	97.3	98.5	105.6	100.5	90.2	95.4	86.7	108.1	96.0	95.6	85.5
22	98.3	100.7	96.2	98.4	100.4	98.2	104.8	100.1	91.5	95.2	86.3	111.1	96.1	95.8	84.4
29	98.3	100.5	96.2	97.9	101.1	98.3	103.5	100.4	92.6	95.3	85.8	112.0	96.1	95.6	83.2
Apr. 5	97.9	99.9	95.8	97.5	100.5	98.0	102.9	100.2	93.8	95.2	85.4	111.7	96.4	95.6	81.7
12	97.3	99.8	94.9	98.3	95.9	97.3	104.0	99.2	95.0	95.2	85.0	107.0	96.4	95.6	81.9
19	96.7	99.3	94.3	98.0	93.9	96.8	103.3	98.5	95.3	95.4	84.6	105.1	96.5	95.5	80.9
26	96.7	99.3	94.4	97.4	93.8	97.3	103.4	98.7	96.4	96.0	83.2	104.8	96.6	95.4	83.7
May 3	96.7	100.2	93.9	98.1	93.4	96.9	105.1	99.6	97.4	94.4	82.7	104.4	96.7	95.0	83.1
10	95.7	99.2	93.1	96.3	93.8	96.2	104.2	98.0	96.6	93.9	80.3	104.4	96.8	95.2	82.8
17	95.9	98.1	94.2	95.6	93.3	97.3	102.8	97.2	96.8	94.0	79.9	104.2	96.8	94.6	92.1
24	95.9	98.0	94.3	95.3	93.1	97.5	102.4	97.1	98.9	93.7	79.6	104.0	97.2	94.3	93.1
31	95.6	97.6	94.3	94.9	93.1	97.3	101.9	96.5	101.4	93.2	79.2	103.9	97.3	94.0	92.7
June 7	95.8	98.5	94.2	95.4	93.3	97.3	103.2	96.6	103.0	92.6	78.8	104.1	97.4	93.6	92.5
14	96.6	99.2	95.2	96.3	94.3	97.8	103.3	97.2	110.2	93.5	79.8	104.0	97.6	93.5	92.3
21	97.5	101.2	95.5	97.9	95.8	98.0	106.3	98.8	115.3	93.0	79.9	104.2	97.7	93.4	91.3
28	97.7	101.6	95.5	97.4	96.2	98.6	105.2	100.1	116.9	92.8	80.0	104.2	97.7	93.3	90.4
July 5	98.1	102.8	95.5	97.6	97.0	99.1	106.0	100.5	117.4	92.8	80.1	104.3	97.9	93.5	89.6
12	98.0	103.4	94.9	96.8	97.2	99.6	106.5	100.6	117.6	92.6	78.5	104.1	98.0	93.3	88.8
19	98.8	105.7	94.5	97.6	97.8	100.3	109.4	103.8	117.8	92.9	77.4	104.5	98.3	92.8	88.1
26	99.1	106.7	94.6	98.1	98.2	100.5	110.2	104.8	117.8	93.0	77.7	104.7	98.2	93.4	87.3
Aug. 2	98.5	105.6	94.3	97.4	98.2	99.8	108.6	104.1	117.6	93.5	78.1	104.8	98.3	93.5	83.8
9	97.6	104.4	93.6	95.9	98.8	98.8	107.7	102.6	118.3	93.6	76.3	104.8	97.3	92.9	82.9
16	97.2	103.5	93.6	95.9	98.0	98.2	107.8	101.6	118.3	94.5	76.6	104.9	97.0	92.7	82.1
23	97.0	103.5	93.3	95.8	97.9	98.1	108.5	102.5	118.8	94.7	75.3	104.9	97.0	92.5	82.0
30	96.9	103.0	93.4	96.3	97.5+	97.4	108.9	101.3	119.6	94.7	75.6	104.7	96.8	92.6	81.3
Sept. 6	96.3	102.5	92.6	96.6	98.1	95.7	109.2	100.9	120.3	94.5–	73.5–	104.7	96.4	92.8	80.7

13	96.3	102.2	92.7	96.5	98.6	95.6	108.4	101.2	120.6	94.7	73.7	104.7	96.7	92.8	80.0
20	96.1	102.3	92.3	97.0	99.3	94.7	109.1	102.1	121.6	94.6	73.9	104.6	96.6	93.0	75.9
27	95.8	101.8	92.2	97.8	98.8	93.7	110.3	101.3	119.0	94.6	74.3	104.2	96.3	93.4	75.8
Oct. 4	94.7	100.0	91.6	96.4	98.8	92.6	108.7	99.7	116.9	94.3	73.3	103.9	96.3	93.7	75.6
11	94.7	99.5	92.1	97.1	98.5	92.0	107.9	99.0	113.4	93.8	75.9	103.4	96.5	93.8	75.8
18	94.6	99.5	92.1	97.4	97.9	91.6	108.3	99.1	113.0	93.4	76.3	103.2	96.3	93.9	75.7
25	94.1	98.7	92.2	97.0	97.8	91.2	107.2	98.3	112.6	93.0	76.7	103.1	95.9	93.9	76.0
Nov. 1	93.7	97.9	92.2	96.2	97.0	91.1	105.8	97.5	111.7	92.3	77.1	102.8	95.8	93.7	76.3
8	93.3	97.6	92.0	95.6	96.4	90.9	105.3	97.0	108.3	91.8	77.5	102.3	96.1	93.8	76.1
15	92.6	97.0	91.6	94.9	95.0	90.6	104.8	95.9	100.9	91.4	77.9	101.8	95.1	93.7	77.5
22	92.2	97.2	91.1	93.9	93.8	90.9	104.2	96.2	97.0	90.8	77.0	101.8	95.3	93.8	77.8
29	92.3	97.5	91.2	94.0	93.9	90.9	103.9	96.1	98.1	90.5	77.4	102.0	95.3	94.0	77.3
Dec. 6	92.7	98.6	91.3	94.3	93.5	91.7	104.0	97.1	99.0	90.3	77.8	101.8	95.5	93.8	77.7
13	92.6	98.0	91.5	94.0	93.6	91.6	103.9	95.9	99.1	89.6	78.3	101.8	97.1	93.9	77.9
20	92.8	98.8	91.2	94.9	92.2	91.7	105.4	96.4	95.0	89.1	78.7	101.7	97.1	93.8	78.3
27	93.1	99.5	91.3	95.0	92.1	92.2	105.3	97.2	95.4	89.1	79.1	101.4	97.2	93.8	78.6
MONTHLY															
Jan.	97.4	98.0	97.5	97.1	99.8	95.5	102.7	99.2	104.8	97.6	84.6	105.0	97.0	94.3	88.6
Feb.	97.7	100.3	96.4	97.2	99.3	97.5	103.9	100.7	96.2	94.9	85.8	106.2	96.8	95.4	89.7
Mar.	98.3	101.1	96.0	98.5	99.1	98.3	105.1	100.5	91.3	95.2	86.5	109.4	96.1	95.6	85.2
Apr.	97.1	99.6	94.7	97.8	95.7	97.3	103.6	99.1	95.4	95.3	84.4	106.8	96.5	95.5	82.1
May	95.9	98.5	93.9	95.9	93.4	97.0	103.1	97.5	98.2	93.8	80.1	104.2	97.0	94.6	89.2
June	96.9	100.2	95.1	96.8	94.9	97.9	104.5	98.2	111.4	93.0	79.7	104.1	97.6	93.5	91.6
July	98.5	104.9	94.8	97.5	97.6	99.9	108.2	102.8	117.7	92.9	78.3	104.5	98.1	93.3	87.7
Aug.	97.3	103.7	93.5+	96.1	98.1	98.2	108.3	102.1	118.8	94.3	76.1	104.8	97.1	92.7	82.2
Sept.	96.0	102.0	92.4	97.0	98.8	94.7	109.2	101.3	120.1	94.6	73.8	104.5–	96.5–	93.1	77.7
Oct.	94.4	99.1	92.1	96.8	98.0	91.7	107.6	98.7	113.4	93.3	76.0	103.3	96.2	93.8	75.9
Nov.	92.7	97.3	91.5	94.7	94.9	90.8	104.6	96.4	101.8	91.2	77.4	102.0	95.5	93.8	77.1
Dec.	92.8	98.7	91.3	94.5	92.8	91.8	104.6	96.6	97.1	89.5	78.6	101.7	96.8	93.8	78.2
QUARTERLY															
1st quarter	97.8	99.8	96.6	97.6	99.4	97.1	103.9	100.1	97.4	95.9	85.6	106.9	96.6	95.1	87.5
2nd quarter	96.6	99.4	94.6	93.5	94.7	97.4	103.7	98.3	101.7	94.0	81.4	105.0	97.0	94.5	87.7
3rd quarter	97.3	103.6	93.6	96.9	98.2	97.6	108.6	102.2	118.9	93.9	76.1	104.6	97.2	93.0	82.5
4th quarter	93.3	98.4	91.7	95.4	95.2	91.5	105.6	97.2	104.1	91.4	77.3	102.3	96.1	93.8	77.1
ANNUAL															
1929	96.3	100.3	94.1	96.7	96.9	95.9	105.5	99.4	105.5	93.8	80.1	104.7	96.8	94.1	83.8

TABLE IV

WHOLESALE COMMODITY PRICE INDEXES (1926=100)

1930

WEEKLY

Week ending	All products (120)	Agricultural products (30)	Non-agricultural products (90)	Raw materials (34)	Semi-mfd. goods (22)	Finished goods (64)	Farm products (15)	Foodstuffs (19)	Hides and leather (8)	Textiles (16)	Fuel and light (10)	Metal products (18)	Building material (20)	Chemicals (14)	Miscellaneous (7)
Jan. 3	93.0	98.5	91.5	94.6	92.2	92.3	104.1	96.6	96.3	89.5	79.6	101.6	97.2	93.3	79.0
10	93.1	98.2	91.7	94.6	92.2	92.5	103.8	95.8	97.5	89.2	80.2	101.3	97.3	93.4	79.4
17	93.2	98.6	91.5	94.3	92.0	93.0	102.5	97.0	98.9	89.2	80.7	99.9	96.5	93.4	79.5
24	93.7	100.4	91.3	94.3	91.7	94.3	102.1	99.4	99.9	88.9	80.5	99.8	96.4	93.5	79.6
31	93.3	99.9	90.9	93.0	91.7	94.2	100.1	99.2	97.1	87.9	80.6	99.9	96.1	93.4	79.8
Feb. 7	93.4	100.5	90.7	92.7	91.5	94.7	99.8	100.0	96.6	87.7	80.8	99.9	96.7	93.3	79.4
14	93.1	100.6	90.0	92.9	91.4	94.1	100.2	100.5	96.2	87.7	80.8	99.8	96.6	93.3	75.3
21	92.2	99.2	89.3	90.8	91.0	94.1	97.8	99.7	95.1	87.6	79.0	99.8	96.4	93.3	75.2
28	91.9	98.7	88.9	90.8	90.2	93.6	97.8	98.9	95.8	86.2	78.9	99.7	96.3	93.0	74.9
Mar. 7	91.0	96.8	88.5	89.4	89.6	93.0	94.8	96.8	96.4	84.5	78.9	99.3	95.9	92.3	74.8
14	90.9	96.5	88.3	89.0	89.5	93.1	94.1	97.2	97.0	83.5	79.1	99.1	95.5	92.0	74.6
21	90.7	95.9	88.4	88.2	89.8	92.8	93.6	96.6	97.7	83.6	79.0	99.3	95.4	91.8	74.8
28	90.4	95.1	88.5	88.8	90.1	92.2	93.4	95.1	98.4	83.5	78.9	99.7	95.4	91.7	74.5
Apr. 4	90.8	95.8	88.6	89.6	90.4	92.2	95.0	95.4	99.1	84.0	78.8	99.4	95.1	91.6	75.0
11	91.3	96.4	89.0	90.2	90.0	92.8	96.0	96.0	99.8	84.0	79.6	99.4	95.1	91.7	75.2
18	90.7	95.9	88.1	90.0	86.0	92.8	95.6	96.3	99.8	83.7	79.6	95.5	95.0	91.4	75.1
25	90.2	95.1	87.8	89.3	85.4	92.5	94.9	95.7	99.5	83.4	79.4	95.3	93.9	91.0	75.3
May 2	89.7	94.2	87.5	88.9	84.8	91.8	94.3	95.2	100.1	83.5	78.3	95.0	93.5	90.6	75.1
9	89.0	94.4	86.1	88.3	82.6	91.4	95.3	94.7	99.5	83.5	75.9	92.8	92.6	90.1	75.3
16	88.7	93.4	86.3	87.7	83.2	91.1	94.4	93.9	99.8	83.5	75.6	93.5	92.7	90.3	75.6
23	88.4	93.2	85.6	87.0	83.1	90.9	94.9	93.4	101.1	83.4	73.4	93.3	92.4	90.3	75.9
30	87.8	91.6	85.6	86.0	82.8	90.5	93.1	92.1	102.6	83.4	73.3	93.4	92.3	90.4	74.8
June 6	87.6	91.2	85.4	85.8	83.0	90.2	92.7	92.1	103.2	83.5	73.2	93.4	91.7	90.3	74.0
13	87.0	90.0	85.1	84.8	82.0	90.0	91.0	92.3	103.0	82.8	73.2	93.0	91.8	89.9	73.2
20	86.2	88.8	84.6	83.2	81.3	89.9	89.1	92.7	102.1	81.4	72.6	92.7	91.7	89.6	73.4
27	85.7	88.0	84.2	83.0	81.6	88.8	88.8	92.2	104.8	80.7	72.8	92.0	91.4	89.1	71.0
July 4	85.6	87.6	84.2	82.9	80.9	88.8	88.8	91.9	103.4	81.0	73.2	91.6	91.4	88.9	71.1
11	84.5	86.6	83.0	80.6	80.6	88.3	86.8	91.6	101.9	80.1	71.5	91.6	91.4	88.5	68.4
18	83.4	85.5	82.1	79.1	80.3	87.4	84.8	90.4	100.6	80.0	70.8	91.3	91.3	88.5	65.0
25	83.3	85.2	82.1	78.9	80.2	87.2	84.4	89.9	99.7	79.9	71.0	91.4	91.0	88.0	65.0
Aug. 1	82.9	84.8	81.9	79.0	79.7	86.4	84.0	89.3	97.9	79.6	72.3	91.6	90.4	87.8	62.2
8	83.1	85.4	81.7	78.9	79.3	86.9	84.3	88.6	96.5	80.7	71.9	91.9	90.0	88.0	61.3
15	83.6	86.8	81.8	80.1	78.9	87.4	86.4	90.9	95.3	80.8	72.1	91.8	89.8	87.6	62.3
22	82.8	85.8	80.8	79.6	78.2	85.9	85.5	89.7	95.3	75.1	72.6	91.4	89.3	87.3	61.9
29	83.3	86.5	81.3	80.9	78.4	85.9	86.5	89.3	95.0	75.3	74.1	91.4	88.8	87.1	62.1

MONTHLY																
Jan.	93.3	99.2	91.3	94.1	91.9	93.4	102.3	97.7	98.2	88.8	80.4	100.3	96.6	93.4	79.5
Feb.	92.7	99.7	89.7	91.8	91.0	94.1	98.9	99.8	95.9	87.3	79.9	99.8	96.5	93.2	76.3
Mar.	90.8	96.1	88.4	88.9	89.8	92.8	94.0	96.3	97.5	83.8	78.9	99.4	95.5	91.9	74.7
Apr.	90.6	95.6	88.2	89.6	87.4	92.5	95.2	95.8	99.7	83.7	79.2	97.0	94.6	91.3	75.1
May	88.6	93.3	86.1	87.5	83.1	91.1	94.5	93.7	100.6	83.5	74.9	93.4	92.6	90.3	75.4
June	86.4	89.1	84.7	84.0	81.8	89.6	90.1	92.3	103.3	81.9	73.0	92.6	91.6	89.6	72.6
July	83.7	85.7	82.5	79.8	80.3	87.5	85.4	90.4	100.4	80.0	71.6	91.5	91.1	88.3	65.7
Aug.	83.2	86.1	81.3	79.9	78.7	86.4	85.7	89.6	95.6	77.3	72.7	91.6	89.5	87.5	61.9
QUARTERLY																
1st quarter	92.3	98.3	89.8	91.6	90.9	93.4	98.4	97.9	97.2	86.6	79.7	99.8	96.2	92.8	76.8
2nd quarter	88.5	92.7	86.3	87.0	84.1	91.0	93.3	93.9	101.2	83.0	75.7	94.3	92.0	90.4	74.4

NOMOGRAMS FOR CORRECTING SIMPLE AND MULTIPLE CORRELATION COEFFICIENTS

By Harold D. Griffin, *Crescent College*

I

A zero-order correlation coefficient frequently requires correction for errors of observation. If the number of observations is large these errors tend to cancel one another in so far as they affect the mean. But the standard deviation is a squared concept, hence the fluctuations due to these errors, being positive, cumulate and the sigmas tend to be larger than they should. And as these sigmas form the denominator in the correlation equation, $r_{xy} = p_{xy}/\sigma_x \sigma_y$, the coefficient of correlation is decreased, or "attenuated."

If we are using Spearman's full formula for attenuation it will be necessary to make two independent measures of each capacity and obtain the reliability coefficients of each test. Thus:

$$\bar{r}_{xy} = (r_{x_1y_1} r_{x_2y_2})^{1/2} / (r_{x_1x_2} r_{y_1y_2})^{1/2}, \text{ where}$$

x_1 and y_1 = the first series of scores obtained in x and y respectively,
x_2 and y_2 = the second series of scores obtained in x and y respectively,
$\bar{r}_{xy}$ = the coefficient of correlation between x and y corrected for attenuation,
$r_{x_1y_1}$ and $r_{x_2y_1}$ = the obtained correlation between x_1 and y_1, and x_2 and y_1 respectively, and
$r_{x_1x_2}$ and $r_{y_1y_2}$ = the reliability coefficients of each test.

The procedure may be shortened by correlating halves of the tests rather than administering the tests twice to the same group. In such a case the formula, $r = 2r/(1+r)$, should be used in obtaining

$$r_{x_1x_2},\ r_{y_1y_2},\ r_{x_1y_2}, \text{ and } r_{x_2y_1}.$$

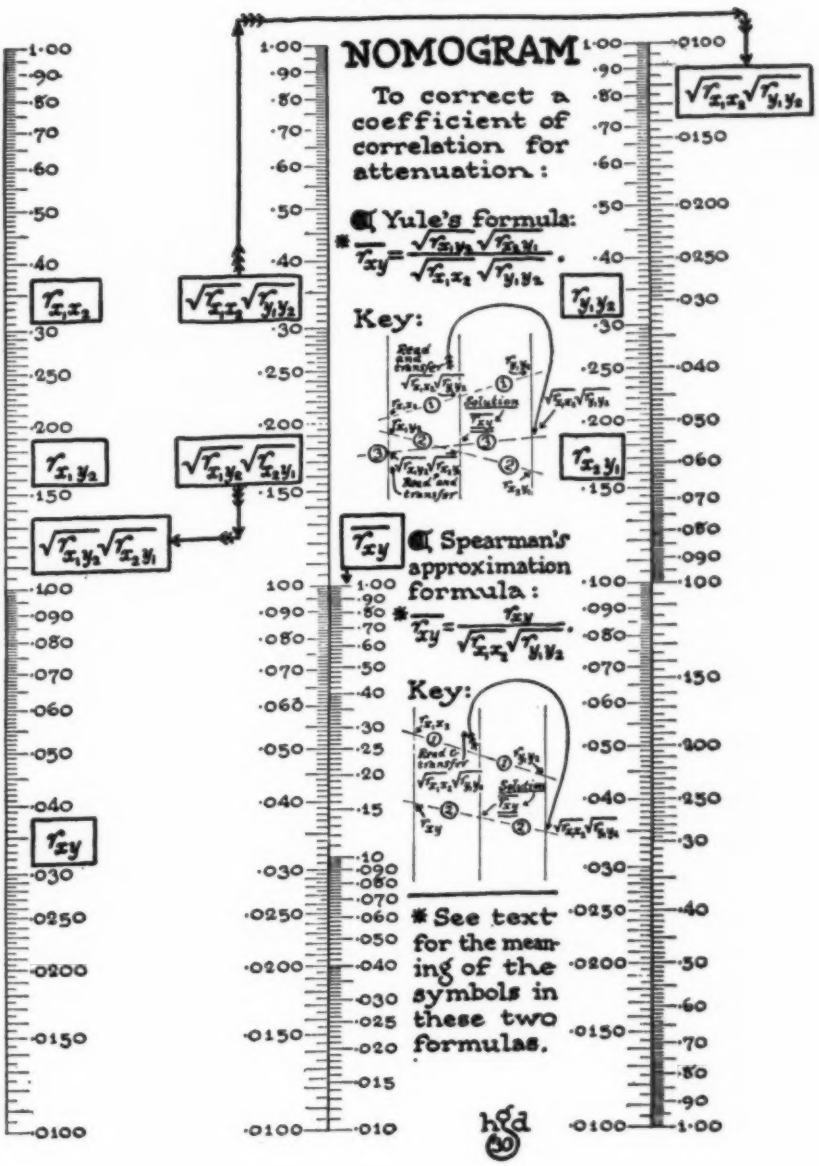
A still further shortening will give an approximate correction by Spearman's approximation formula:

$$\bar{r}_{xy} = r_{xy} / (r_{x_1x_2} r_{y_1y_2})^{1/2}, \text{ where}$$

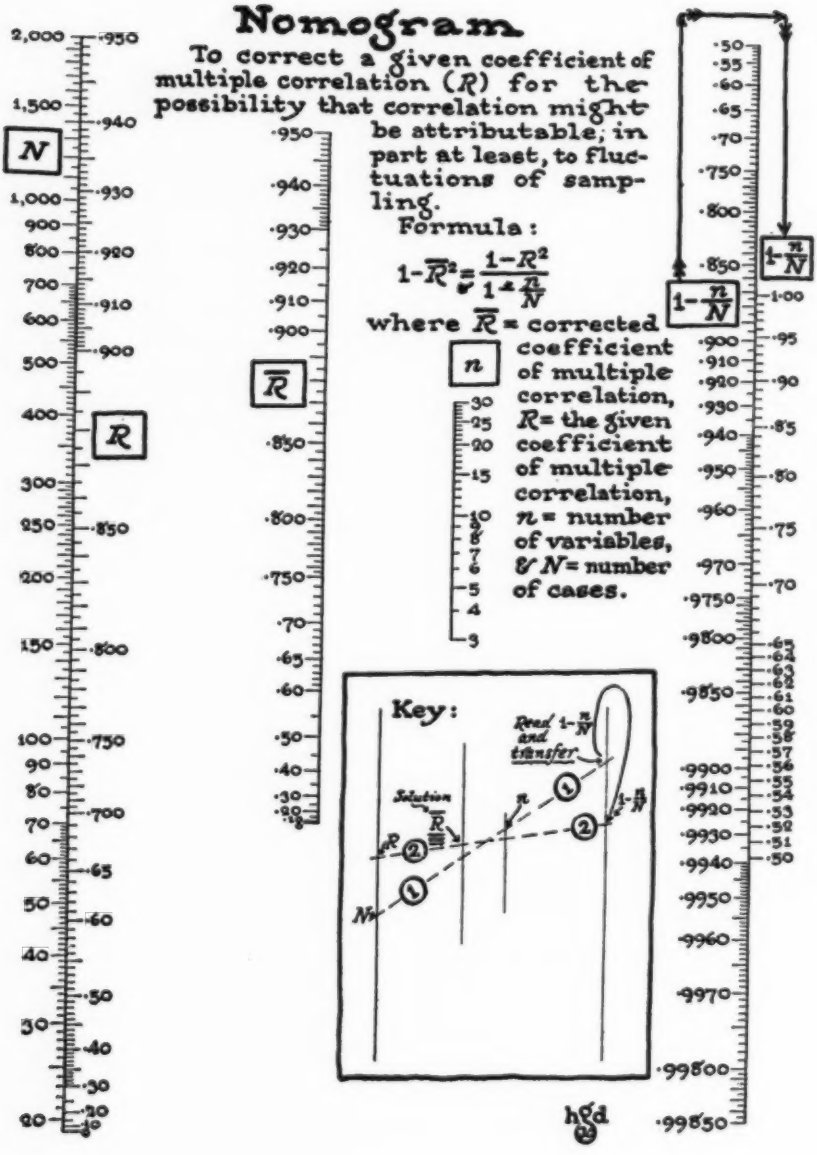
r_{xy} = the zero-order correlation between x and y, and the other symbols have the same meanings as before.

Nomogram I will enable the reader to correct a coefficient of correlation for attenuation either by Spearman's full formula or by his approximation formula. The calculations may be done with a straight-

NOMOGRAM I


NOMOGRAM
To correct a coefficient of correlation for attenuation:
Yule's formula:
Key:
Spearman's approximation formula:
Key:
* See text for the meaning of the symbols in these two formulas.
hgd
30

NOMOGRAM II


Nomogram
To correct a given coefficient of multiple correlation (R) for the possibility that correlation might be attributable, in part at least, to fluctuations of sampling.
Formula:
$1-\overline{R}^2=\frac{1-R^2}{1-\frac{n}{N}}$
where $\overline{R}$ = corrected coefficient of multiple correlation, R = the given coefficient of multiple correlation, n = number of variables, & N = number of cases.
N
R
$\overline{R}$
n
$1-\frac{n}{N}$
Key:
Read and transfer
Solution
hgd

edge or taut thread, but the most satisfactory way is to make an isopleth by scratching a straight line with a needle on the underside of a strip of celluloid of good weight.

II

The coefficient of multiple correlation is also subject to errors arising from the fact that it is a positive concept which increases with the addition of each new variable—although this increase may be very small. Therefore it is possible that fluctuations due to sampling have played a part in determining its size.

B. B. Smith has evolved a formula to correct a coefficient of multiple correlation for fluctuations due to sampling.[1] This formula is

$$\overline{R} = [1-(1-R^2)/(1-n/N)]^{\frac{1}{2}}, \text{ where}$$

$\overline{R}$ = the corrected multiple correlation coefficient,
R = the raw multiple correlation coefficient,
n = the number of variables, and
N = the number of cases.

Nomogram II is designed to calculate this formula with the minimum of labor.

[1] B. B. Smith, "Forecasting the Acreage of Cotton," this JOURNAL, 20 (1925): 31–47, p. 41, note; also B. B. Smith, "Another Attempt to Explain Mu'tiple Correlation in Simple Terms," this JOURNAL, 24 (1929): 61–5.

ON THE THEORY OF CONTINGENCY

I

NOTE ON PROFESSOR J. ARTHUR HARRIS' PAPERS[1] ON THE LIMITATION IN THE APPLICABILITY OF THE CONTINGENCY COEFFICIENT

By Karl Pearson

I venture to think that my friend Professor Harris has overlooked the essential nature of contingency. It consists in putting into each cell the number of individuals to be expected there, if the variates were *unrelated*. Cells wherein the existence of individuals is arithmetically, physically or physiologically impossible are simply to be disregarded. As Professor Harris remarks, contingency is independent of the order of the columns or rows. It is shortest therefore to write Professor Harris' Table I as follows:

TABLE I

COEFFICIENT OF ASYMMETRY

Locular Composition	.0000	.9428	1.6330	1.8856	.4714	.8165	1.2472	1.4142	1.6997	Totals
3 even 0 odd...	462	130	2	1	...	...	..	..	..	595
0 even 3 odd...	103	35	1	0	...	...	..	..	..	139
2 even 1 odd...	...	...	..	..	614	138	21	14	1	788
1 even 2 odd...	...	...	..	..	443	95	22	8	5	573
Totals.....	565	165	3	1	1,057	233	43	22	6	2,095

Looked at in this form it is seen that the right-hand upper and left-hand lower quadrants are areas of impossible occurrence. The sample may be supposed extracted from a population of the following kind:

TABLE II

π_{11} π_{21}	π_{12} π_{22}	π_{13} π_{23}	π_{14} π_{24}	... ...	... ...	... ...	... ...	... ...	$\pi_{1.}$ $\pi_{2.}$
... ...	... ...	... ...	... ...	π_{15} π_{25}	π_{16} π_{26}	π_{17} π_{27}	π_{18} π_{28}	π_{19} π_{29}	$\pi_{3.}$ $\pi_{4.}$
$\pi_{.1}$	$\pi_{.2}$	$\pi_{.3}$	$\pi_{.4}$	$\pi_{.5}$	$\pi_{.6}$	$\pi_{.7}$	$\pi_{.8}$	$\pi_{.9}$	M

Here it is quite impossible for individuals to occur in the cells occupied by dots. We are really therefore dealing with two independent distributions.

[1] This Journal, Vol. 22, pp. 460–72, and Vol. 24, pp. 367–75.

Let $M_{\cdot 1} = m_{\cdot 1} + m_{\cdot 2} + m_{\cdot 3} + m_{\cdot 4}$
and $M_{\cdot 2} = m_{\cdot 5} + m_{\cdot 6} + m_{\cdot 7} + m_{\cdot 8} + m_{\cdot 9}$
$M_{1\cdot} = m_{1\cdot} + m_{2\cdot}$.
$M_{2\cdot} = m_{3\cdot} + m_{4\cdot}$.

and of necessity $M_{\cdot 1} = M_{1\cdot}$ and $M_{\cdot 2} = M_{2\cdot}$.
For independence:

$$m_{11} = \frac{m_{\cdot 1} m_{1\cdot}}{M_{1\cdot}} = \frac{m_{1\cdot} m_{\cdot 1}}{M_{\cdot 1}}$$

$$m_{35} = \frac{m_{3\cdot} m_{\cdot 5}}{M_{\cdot 2}} = \frac{m_{\cdot 5} m_{3\cdot}}{M_{2\cdot}} \text{ and so on.}$$

Now if m_{st} be the contents of the s, t th of the *possible* cells of the sampled population and n_{st} of the corresponding cell of the sample we have for the mean square contingency

$$\phi^2 = S_{s,t}\left[\frac{\left(m_{st}\frac{N}{M} - n_{st}\right)^2}{N\left(m_{st}\frac{N}{M}\right)}\right],$$

$$= S_{s,t}\left[\frac{(\overline{n}_{st} - n_{st})^2}{N n_{st}}\right],$$

where $\overline{n}_{st} = \frac{N}{M} m_{st}$.

Thus $$\overline{n}_{11} = \frac{N}{M}\frac{m_{\cdot 1} m_{1\cdot}}{M_{1\cdot}} = \frac{\frac{N}{M} m_{\cdot 1} \times \frac{N}{M} m_{1\cdot}}{\frac{N}{M} M_{1\cdot}}$$

$$\overline{n}_{35} = \frac{N}{M}\frac{m_{3\cdot} m_{\cdot 5}}{M_{\cdot 2}} = \frac{\frac{N}{M} m_{3\cdot} \times \frac{N}{M} m_{\cdot 5}}{\frac{N}{M} M_{\cdot 2}} \text{ and so on.}$$

But we do not know the sampled population values, so we are compelled to insert the observed values, *i. e.* take

$$n_{11} = \frac{n_{\cdot 1} n_{1\cdot}}{N_{1\cdot}}$$

$$n_{35} = \frac{n_{3\cdot} n_{\cdot 5}}{N_{\cdot 2}} \text{ and so forth.}$$

This is the usual custom with the computation of ϕ^2. In our case

$$N_{1.} = n_{1.} + n_{2.} = 595 + 139 = 734 = N_{.1}$$

$$N_{.2} = n_{.5} + n_{.6} + n_{.7} + n_{.8} + n_{.9} = 1361 = N_{.2}.$$

Observing these relations we may now put Professor Harris' Table I into a form showing the expected results on the basis of no contingency or association between his variates into the form of Table III where beneath each of his numbers is the number to be expected on this hypothesis. It will be seen that the expected numbers are very close to the observed, or there is very little contingency.

TABLE III

Actual	462	130	2	1	...	...	..	..	..	595
Expected if no contingency	458	134	2	1						
Actual	103	35	1	0	...	...	..	..	..	139
Expected if no contingency	107	31	1	0						
Actual	...	..	..	..	614	138	21	14	1	788
Expected if no contingency					612	135	25	13	3	
Actual	...	..	..	..	443	95	22	8	5	573
Expected if no contingency					445	98	18	9	3	
	565	165	3	1	1,057	233	43	22	6	2,095

For this table, I have given the expected only their nearest whole numbers, and this I take it is *theoretically* correct.

We have:

$$\phi^2 = .0026{,}0250$$

and $C_2 = .0509$.

If we take the standard error for zero contingency, approximately equal to

$$\sigma_{\phi^2} = \frac{\sqrt{2(c-1)}}{N}$$

where c is the number of cells, we should have in this case

$$\sigma_{\phi^2} = .0027{,}8327$$

or ϕ_1 is less than its standard error. It seems highly probable, therefore, that there is no association whatever in Professor Harris' data between the Coefficient of Asymmetry and the Locular Composition.

It may be said that this result is inconsistent with Professor Harris' values obtained by other methods. I am not surprised at it, because I do not think those methods are at all applicable. There is no real order

or continuity in Locular Composition, and no continuity in the Coefficient of Asymmetry. I believe that the coefficient of mean squared contingency is the *only* suitable method of measuring the association, but it must be dealt with from the fundamental principle of contingency, *i. e.* writing in each cell the probable frequency which would occur there if there was no association; these frequencies must be computed *only* for cells where the occurrence of frequency is possible, and this denotes a careful *a priori* consideration of how the possible arises.

The remarks made here apply also to Professor Harris' first paper. His Table II therein is precisely the case dealt with in *Biometrika* twelve years before the publication in this JOURNAL of Professor Harris' difficulty as to the contingency coefficient leading to an erroneous value. That misleading value is again a case of Professor Harris' putting contingency values into cells where they ought not to be. If he examines *Biometrika*, Volume X, 1914, pages 434 to 445, in the article by Dr. Waite, he will find the treatment of a number of such coefficients of contingency for what I have termed "restricted tables," *i. e.* tables in which certain cells cannot contain frequency. They occur in the association of finger-print patterns on the same or both hands, where the number of digits and therefore of finger prints is limited. It is, perhaps, needless to add that these restricted tables occur also in the correlation between the numbers of cards of the same suit in the players' hands at whist, etc.

My excuse for this note must be that tables presenting the same difficulties as those noted by Professor Harris were recognized many years ago, and I do not think they form any real limitation to the use of contingency methods. That C_2 will not be the coefficient of correlation for such tables is clear, for usually the correlation coefficient has no meaning—the regression is of necessity skew. For this see the regression curves in Dr. Waite's paper, Plates $\alpha - \epsilon$.

II

PROFESSOR PEARSON'S NOTE ON OUR PAPERS ON CONTINGENCY

BY J. ARTHUR HARRIS, ALAN E. TRELOAR AND MARIAN WILDER

Notwithstanding the precision of statement which those who work in the statistical field seek to attain, misunderstandings in regard to matters of detail must occasionally arise, even when they are in full agreement on essentials. This has been the case in Professor Karl Pearson's criticism of our papers on contingency.

Professor Pearson's note has indicated in our work only one omission for which we are glad to have the opportunity to apologize. He is quite right in calling attention to our neglect of his own work on "restricted tables" published a number of years ago. We regret this omission and have no desire to seek exoneration for our oversight either on the ground that the work appeared incidentally in a paper on finger prints by Waite [1] or on the ground that the case with which he dealt seems to us quite different from either of the two with which we were concerned.

As far as the existence of any difference of opinion on larger essentials between Professor Pearson and ourselves is concerned, this note might well close with our apology for having overlooked Pearson's treatment of this special finger-print problem. Since, however, some of Professor Pearson's statements might lead the casual reader to feel that there is dissention in the biometric camp, or to conclude that if biometricians cannot agree among themselves their work may be disregarded by others, it is worth while to show how exactly in agreement we are.

(1) In his opening sentence Professor Pearson says, "I venture to think that my friend Professor Harris has overlooked the essential nature of contingency." Again, in referring to our first paper,[2] he says: "That misleading value is again a case of Professor Harris' putting values into cells where they ought not to be." These sentences might be construed by a superficial reader as indicating that we had blundered.

We had the exact Pearsonian conception of contingency so clearly in mind that what we actually did was to show that *if* the classical contingency method of Pearson [3] *be applied without modification to certain types of tables, it will put values into cells where they ought not to be.*[4] We then proceeded to seek for some method of overcoming this limitation of his original method *as applied to such tables.*

We dealt with a correlation surface for quantitatively measured variables because it is only in such cases that the agreement of r, η, and C_1 can be tested. In doing this we had the objective of preventing error on the part of those who might use the contingency method of Pearson in its original form for tables of two non-measurable variables

[1] H. Waite, "Association of Finger Prints," *Biometrika* 10: 421–478, 1915.

[2] J. Arthur Harris and Alan E. Treloar, "On a Limitation in the Applicability of the Contingency Coefficient," this JOURNAL 22: 760–772, 1927.

[3] K. Pearson, "On the Theory of Contingency and its Relation to Association and Normal Correlation," *Drapers' Company Research Memoirs, Biometric Series*, 1, London, 1904.

[4] Professor Pearson's original memoir called for the use of theoretical frequencies determined by *independent* probability. Obviously theoretical frequencies thus calculated for a surface must occur in all cells. The cases presented by ourselves represent real limitations to the theory as originally stated.

and who might inadvertently apply his unmodified theory to cases in which some of the cells were necessarily void.

This is, in principle, in exact agreement with what Professor Pearson in his note maintains should be done.

(2) While we are to blame for being ignorant of the treatment of the case of "restricted tables" in the paper on finger prints by Waite, it seems to us that Professor Pearson is in error in assuming that Table II of our first paper (this JOURNAL, 1927, p. 463) "is precisely the case dealt with in *Biometrika* twelve years before." To us the case treated by Waite and that treated by us seem entirely different. Waite worked with a correlation surface between x and y in which $x+y \overline{\leq} 5$ in every case. We worked with a surface in which $y \underline{\leq} x$ in every case.

We cannot deal adequately with the problem of the differences between these two categories here, but shall hope to discuss certain phases of this problem later. For the present it is sufficient to compare with our own value the constant obtained by the application of Waite's method of obtaining contingency when applied to the case treated by us (this JOURNAL, 1927, pp. 460–472).

The actual frequencies in our Table VI show a remarkable linearity of regression[1] of y (number of seeds) on x (number of ovules). This is also characteristic of many other correlation surfaces of the same kind which we have investigated.[2] It is, therefore, quite legitimate to treat such tables by either the product moment correlation or the correlation ratio method. For the data employed by us r_{xy} and ${}_x\eta_y$ differ by only 0.0002, since $r_{xy}=0.6482 \pm .0023$, ${}_x\eta_y=0.6484 \pm .0023$. When contingency is calculated by our first method of removing theoretical frequencies from cells to which they have been put by the method of Professor Pearson's classical paper, but in which they cannot occur, we find $C_1=0.6548$, differing from r_{xy} by only $+0.0066$ and from ${}_x\eta_y$ by $+0.0064$. If Waite's method be applied to this same table it gives $C_1=+0.1471$, a value in the grossest disagreement with both r and η. Neither time nor space permit an adequate treatment of the problem here. The single illustration is sufficient to show that Waite's method is altogether inadequate for a satisfactory treatment of the case which we considered in this JOURNAL in 1927. The result is of particular interest since it indicates that instead of two categories of limitations in the applicability of the conventional contingency

[1] J. Arthur Harris, "On the Relationship Between the Number of Ovules Formed and the Number of Seeds Developing in *Cercis*," *Bulletin of the Torrey Botanical Club* 40: 241–256, 1914.

[2] See, for example: J. Arthur Harris, "On the Relationship Between Bilateral Asymmetry and Fertility and Fecundity," *Arch. f. Entwicklungsmech. Org.* 35: 500–522, 1912; J. Arthur Harris, "Further Studies on the Relationship Between Bilateral Asymmetry and Fertility and Fecundity in the Unilocular Fruit," *Genetics* 2: 186–204, 1917; also incidental tests in several other papers.

method there are probably at least three such categories. It is important that these be borne in mind by biological or sociological workers using contingency methods in order that they may not unwittingly reach misleading results.

(3) Turning now to the problem of our second paper,[1] Professor Pearson rearranges our Table I and treats it in his Tables I–III with accompanying equations. For series A he arrives at the constant $C_1 = 0.0509$.

Now the first entry under Method B in our Table VI is $C_1 = 0.0597$. If Professor Pearson had retained six places of decimals (a needless refinement except for the checking of the results of two methods) in the expected values of his Table III he would also have gotten $C_1 = 0.0597$, agreeing exactly with our result. When we carry our calculations only to the nearest whole individual frequency, as Professor Pearson has done, we get $C_1 = 0.0506$ agreeing exactly with his result.[2]

In short Professor Pearson's method is merely our Method B[3] expressed in different notation. It is naturally a source of satisfaction to us to find our results thus confirmed by Professor Pearson's own analysis of the problem.

(4) Finally, to preclude all possibility of misunderstanding, we must again emphasize, as we have in the papers under consideration and in another which deals with the biological aspects of one of these problems,[4] the fact that our work contains no general criticism of the contingency method. It has been a realization of the importance of the contingency conception as developed by Karl Pearson in his

[1] J. Arthur Harris and Chi Tu, "A Second Category of Limitations in the Applicability of the Contingency Method," this JOURNAL 24: 367–375, 1930.

[2] We believe that Professor Pearson's computer has made a slight arithmetical error (altogether unimportant for any purpose except for the checking of results) and that his value should be $C_1 = 0.0506$, instead of $C_1 = 0.0509$.

[3] While it has no bearing on the problem of the present discussion we may take this occasion to note a misprint in our treatment of Method A (this JOURNAL 24: 367–375).

The formula at the bottom of page 370 should obviously read

$$n''_{\nu} = n'_{\nu} \frac{N}{S_o(n'_{\nu})}$$

instead of

$$n''_{n\nu} = n'_{n\nu} \frac{N}{S_{\nu}(n'_{n\nu})}$$

where S_o denotes summation for the theoretically occupied and S_ν denotes summation for the theoretically void cells. Lines 2 and 3 of p. 371 should read ". . . and S_o denotes summation for the theoretically occupied cells of the whole surface." The formula is correctly given in a different form at the head of the final column of Table VI.

A typographical error of no practical importance occurs in the last entry of the last column of Table VI, which should read .0715 instead of .0745.

[4] J. Arthur Harris, Chi Tu and Marian Wilder "The Biological Significance of Certain Differences Between the Values of the Correlation Coefficient, Correlation Ratio and Contingency Coefficient," *American Journal of Botany* 17: 175–183, 1930.

masterly paper of 1914 that has led us to seek to ascertain and to overcome limitations which may apply in certain special cases. We still feel that (in the words of Karl Pearson): "The degree of approach of both C_1 and C_2 to the correlation must be studied for each special class of cases, and only when this has been done will their use be legitimate and effective."

III

POSTSCRIPT

Owing to the courtesy of Professor Harris I have seen a type-script of his rejoinder to my *Note*. I do not think it needful to modify the latter. I am at the present time away from home and unable to refer to his data in Volume 22 of this JOURNAL, but I will do so on my return to London. I may say, however, that the conditions $x+y \lesseqgtr 5$ and $y \lesseqgtr x$ are *mathematically* of the same character, *i. e.* the area of impossible cell content is in both cases bounded by a straight line. Whether there is any other restricting condition in Professor Harris' case I cannot say without studying his data.

I would remark, however, that the equality of the coefficient of mean square contingency with the coefficient of correlation and the correlation ratio, only follows when (i) the distribution is truly normal, and (ii) when the three coefficients have been corrected for grouping. It does not depend solely on the linearity of regression, but on the homoscedasticity of the arrays. True normality of distribution is theoretically impossible within an angular area. I would add finally that a general theory of contingency subject to restricting conditions was discussed by me many years ago in a paper entitled "Partial Contingency" in *Biometrika*.

KARL PEARSON

IV

Due to the untimely death of Professor Harris the Postscript must stand unanswered.

THE EDITOR

THE PROGRAM FOR THE 1930 ANNUAL MEETING

By M. C. Rorty

The plans for the annual meeting of the Association to be held at Cleveland from December 29 to December 31, inclusive, have involved certain somewhat novel considerations.

In the first place, the 1930 meeting is to be held in association with the annual meeting of the American Association for the Advancement of Science, and the natural sciences will, therefore, occupy a prominent position in the various discussions and joint meetings, although the American Economic Association, the American Sociological Society, and other associations with which this Association has customarily met in the past, will also hold their annual meetings in Cleveland.

The Cleveland meeting will be one of the largest gatherings of scientific societies held within recent years, since annual meetings will be held simultaneously by the following learned societies and sections of the A. A. A. S.:

Mathematics, Section A, American Mathematical Society, Mathematical Association of America.

Physics, Section B, American Meteorological Society, American Physical Society.

Chemistry, Section C.

Astronomy, Section D.

Geology and Geography, Section E.

Zoölogical Sciences, Section F, American Association of Economic Entomologists, American Society of Parasitologists, American Society of Zoölogists, Entomological Society of America, Wilson Ornithological Club.

Botanical Sciences, Section G, American Phytopathological Society, American Society of Plant Physiologists, Botanical Society of America.

Societies related to both Zoölogical and Botanical Sciences, American Microscopical Society, American Society of Naturalists, Ecological Society of America, Genetics Section of American Society of Zoölogists and Botanical Society of America, Phi Sigma Biological Research Society.

Anthropology, Section H, American Anthropological Association, American Folk-Lore Society.

Psychology, Section I.

Social and Economic Sciences, Section K, American Association for Labor Legislation, American Association of Teachers of Marketing and Advertising, American Association of University Instructors in Accounting, American Economic Association, American Political Science Association, American Sociological Society, American Statistical Association, Farm Economics Association, Metric Association, Stable Money Association.

Historical and Philological Sciences, Section L, History of Science Society.

Engineering, Section M.

Medical Sciences, Section N, American Society of Tropical Medicine.

Agriculture, Section O, American Society of Agronomy, American Society for Horticultural Science, Association of Official Seed Analysts, Geneticists Interested in Agriculture, Potato Association of America.

Education, Section Q, Phi Delta Kappa Fraternity.

As a second consideration affecting the plans for the 1930 meeting, the division of responsibilities among and between the several vice-presidents of the Association has been particularly important. This year, for the first time, the actual program is being arranged in sections by the several vice-presidents, with the president acting, in effect, simply as chairman of the program committee.

Still a third consideration has been the recent rapid spread of statistical technique in all branches of science. This tendency of the statistical element to dominate one science after another, plus the rapid spread of statistics in practical applications, has, in fact, become a source of genuine embarrassment to those who classify themselves primarily as statisticians. For this reason, as a first step in working out the program for this year's meeting, it seemed desirable to set up as clear a picture as possible of the exact field that the Statistical Association should attempt to cover. Discussions along this line among the officers and directors of the Association have led to a substantial agreement that the Statistical Association has:

1. A primary and permanent interest in the development and extension into use of a correct and effective statistical methodology.

2. A direct, and presumably primary and permanent, interest in the adequate development and proper correlation of the statistical work of all governmental and public agencies, beginning with the Federal Census and extending through all categories of public statistics.

3. An indirect interest, in coöperation with the Federal Reserve System, the Department of Commerce, the various trade associations, etc., in the development and proper correlation of economic and social statistics, supplementary to the public statistics, that are compiled by non-public agencies.

4. A direct interest in certain statistical applications which are of particular interest to the members of the Association, or which are not being adequately fostered by other organizations. For example, there fall in this category statistical applications in commerce and industry; general financial, investment, and production statistics, particularly as related to business trends; etc.

5. An indirect interest, in coöperation with other organizations, in the development and correlation of statistical applications in fields not within the immediate purview of this Association. For example,

there fall in this category statistical applications in astronomy and other natural sciences, in engineering, etc.

6. A direct interest in coöperation with international statistical associations, and in the development of adequate international statistics.

The preceding outline of the functions of the Association has been used as a general guide in the working out of the program for the present year. First consideration has been given, however, to the setting up of sections of the program along the lines of the division of responsibilities between the vice-presidents, and with particular reference to securing papers from those who have recently done specially significant original work in the statistical field. An attempt has also been made to give a specially prominent position to statistical methodology and to the relations of statistical work to the natural sciences.

The program as a whole cannot assume definite form until after the present summary goes to press. However, arrangements already made have progressed far enough to make it possible to indicate the general scope of the program and a considerable number of the special papers.

In general, it is expected that each vice-president will present a paper to constitute, in effect, a vice-presidential address, which will ordinarily summarize and outline the program over which he is to preside.

Vice-President A. R. Crathorne, under the new organization, deals with statistical and actuarial methods and technique, and the teaching of statistics. His own paper will be under the heading "Principles of Statistical Methodology." This will be supplemented in a general session by short papers, in such manner as to outline in a non-technical way the entire range of statistical methodology, with particular reference to recent developments. The more technical papers will be presented in supplemental sessions, or in joint session with the American Mathematical Society.

The papers already scheduled by Professor Crathorne are as follows: "Baye's Theorem," E. C. Molina; "Statistics without Probability," E. V. Huntington; "The Use of the Normal Hypothesis," B. H. Camp; "Why Scientists Need Statistical Theory," W. A. Shewhart; a non-technical discussion of some of the recent work in small samples, H. L. Rietz; "Recent Advances in Statistical Theory," E. L. Dodd; and a paper elucidating the theory of homogeneity and stability by Professor Bortkiewiez of the University of Berlin.

Vice-President D. R. Belcher is coöperating with Professor Crathorne in the preparation of this program, and expects to make an individual contribution.

The joint session with the American Mathematical Society and with

Sections A and K of the American Association for the Advancement of Science will cover, among other items, a presentation by Professor Hotelling of some of the results of his own work together with a discussion of the recent methods of R. A. Fisher, and perhaps also papers on the mathematical theory of economics by Professor G. C. Evans of Rice Institute, and on new developments in statistics by Professor Ragnar Frisch of Oslo University.

Vice-President R. H. Coats is concerned, in the new organization, with facts and methods relating to political science, sociology, social welfare, labor problems and vital statistics. Dr. Coats' own paper will deal with enumeration and sampling in the field of the census. In this connection it is the expectation that the natural division of census work between direct enumerations and sampling operations will be fully developed by Dr. Coats and in supplemental papers.

In addition to the discussion of census work, Dr. Coats' section of the program will include a joint session with the American Association for Labor Legislation, at which Miss Mary van Kleeck of the Russell Sage Foundation is arranging for several papers dealing with recent developments in unemployment statistics.

Dr. Stuart A. Rice is also arranging a joint session with the American Sociological Society.

Specific papers for Dr. Coats' program, other than those above indicated, will be presented, in joint session with the American Sociological Society or otherwise, by Professor L. C. Marshall, of the Institute of Law, on statistics in the field of law and of civil and criminal procedure, by Professor Dorothy Thomas on the observability of statistical data along sociological lines, by Dr. E. B. Wilson on studies of case records of psychopathics, and by Dr. H. M. Pollock, who will present, for the Committee on Institutional Statistics, a report dealing with the general problem of welfare and institutional statistics in the United States, with particular reference to mental defectives.

Vice-President E. L. Thorndike, to whom is assigned the consideration of facts and methods relating to anthropology, biometry, psychology, and education, will present his own paper on these subjects, and is also arranging a session to be devoted to methods and standards of statistics of public and private education in the United States.

Vice-President W. Randolph Burgess, who, under the new organization, is concerned with facts bearing on economics and economic theory, is organizing an extensive program, including a review of business and economic conditions in 1930 and forecasts for 1931. Other discussions will concern themselves with statistical studies of the money market and statistical methods in the valuation of securities.

Mr. Carl Snyder, as Chairman of the Committee on Price Indices, is also organizing a session on the revision of price indices, at which short papers will be presented by Professors Copeland and Mills, Dr. Ezekiel, and Dr. Robert B. Burgess. In addition, a joint session with the American Economic Association is planned for a discussion of the 1929–1930 business recession.

Vice-President Bradford B. Smith, whose assignment under the new organization includes facts and methods pertaining primarily to business, is arranging three sessions, one on trends in individual trades and industries, a second on methods of business forecasting, and a third on methods of forecasting security markets.

In addition to the preceding, Dr. E. B. Wilson is arranging a session on statistical applications in the natural sciences. Under this heading papers have already been promised by W. F. G. Swann, Director of the Bartol Research Foundation of the Franklin Institute, on the place of statistics and the field of probability in present-day physics, by Jan Schilt, of the Yale University Observatory, on "Some Astronomical Results of the Application of Statistics," and by W. A. Shewhart, on statistical applications in engineering and industrial research.

Other important items on the program will include a special paper by Dr. Dublin on recent developments in the statistical work of the Metropolitan Life Insurance Company, a joint session with the American Farm Economic Association and a round table discussion on the subject of the relations of the American Statistical Association to work in international statistics. This latter discussion will be under the chairmanship of Dr. E. Dana Durand, who was the delegate of the United States to the International Diplomatic Conference on Economic Statistics held under the auspices of the League of Nations in 1928.

Excellent accommodations for members of the Association have been arranged at the Hotel Statler, where ample meeting rooms will also be provided for three simultaneous sessions, wherever necessary. Joint meetings may, however, be held in some instances in the buildings of the Case School of Applied Sciences, where facilities are to be provided for the meetings of the American Association for the Advancement of Science and certain other societies.

A METHOD OF APPRAISING SHORT-TERM FORECASTS

Now that business organizations are more and more making use of statistical methods to predict demand for their products or services, in both the immediate and remote future, appraisal of the success of this very practical application of statistics is beginning to receive attention. The range of time covered by such predictions may be a few months or a period of years. The purpose of this note is to outline a method of appraising forecasts covering a very short term of months. For illustrative material a group of forecasts of the *cyclical movement* of a series of data which reflects demand for telephone service between telephones within the city of Chicago, and which is called local telephone calls, has been used.

To determine the cyclical movements of the series in the past, monthly seasonal indices were first calculated for the period 1922–1929 by the link-relative method. A straight-line trend was then fitted to the data corrected for seasonal variation and the percentage deviations from normal calculated. The movements of a curve of the percentage deviations from normal are referred to hereafter as cyclical movements of the series, and monthly points on this curve as cyclical points.

The forecasts are made as follows: on the first day of each month the curve of the actual cyclical movements is brought up to date by the addition of th actual cyclical point for the preceding month. The curve is then extended two points in advance to yield a forecast of the cyclical point for the current month and for the month following. The extension of the curve is freehand and the direction given to it is based on the judgment of the forecaster as to the future course of industrial production in the Chicago Region, which is believed to have been the most important influence affecting the cyclical movements of local telephone calls in the past. Sixteen two-month forecasts covering the period February, 1929, to May, 1930, are available for consideration in this note. It is apparent that the forecasts overlap, each month in the period covered having been forecasted twice.

In appraising the forecasts an attempt has been made to determine whether the forecasts were of *any* value rather than *how* valuable they were. A qualitative appraisal requires the formulation of standards of accuracy, in which there is necessarily an element of personal bias. The procedure followed was to compare the accuracy of the forecasts with the accuracy of forecasts in which the factor of judgment was omitted. In accordance with this procedure, comparable forecasts of the cyclical movements of the series have been made by two "blind" or "mechanical" methods which are described below. To avoid confusion the method under appraisal will be termed the "judgment" method and the two mechanical methods the "no change" method and the "projection" method.

The "no change" method of forecasting assumes that the two cyclical points of each forecast will be the same as the most recent actual cyclical point. Each forecast is a horizontal projection of the actual cyclical point for the month preceding the date of the forecast. It is apparent that the "no change" method will result in greater accuracy than the "judgment" method in all cases in which

the latter forecasts incorrectly the general direction ("up" or "down") of the actual cyclical movement.[1] A comparison of the results of the two methods shows that, for the first month of the forecasts, the "judgment" method was more accurate than the "no change" method in ten cases; the two methods were equally accurate in three cases; in three cases the "no change" method was more accurate. For the second month of the forecasts the judgment method was more accurate in twelve cases; in four cases the "no change" method was more accurate.

The "projection" method assumes a continuation of the most recent cyclical movement. In view of the short term of the forecasts and the much longer duration of both the upswings and downswings of the cyclical movements in the past, the "projection" method is a reasonable mechanical method of forecasting. Forecasts were made by this method by projecting for two months in advance a line drawn between the actual cyclical points of the first and second months preceding the month in which each forecast was made. For the first month of the forecasts, the "judgment" method was more accurate than the "projection" method in eleven cases; the two methods were equally accurate in two cases; in three cases the "projection" method was more accurate. For the second month of the forecasts, the "judgment" method was more accurate in ten cases; the two methods were equally accurate in three cases; in three cases the "projection" method was more accurate.

It is of interest also to compare the average differences in percentage deviations between forecasts made by the three methods and actual cyclical points.

AVERAGE DIFFERENCES OF FORECASTED CYCLICAL POINTS FROM ACTUAL CYCLICAL POINTS

(Arithmetic Average of Differences, Disregarding Signs)

	First Month of Forecasts	*Second Month of Forecasts*
"Judgment" Method.....	0.7 percentage deviations	1.0 percentage deviations
"No Change" Method....	1.0 " "	1.5 " "
"Projection" Method.....	1.2 " "	2.3 " "

The procedure of appraising short-term forecasts outlined above does not attempt to cope with the problem of qualitative appraisal. It is, however, subject to no personal bias and does serve to indicate whether the factor of judgment has increased or decreased the accuracy of the forecasts.

ROBERT B. KING

AWARDS OF THE SOCIAL SCIENCE RESEARCH COUNCIL

During 1929–30 the Social Science Research Council awarded 41 grants-in-aid from over 100 applications. The total amount involved in these grants is approximately $25,000. The grants-in-aid are open to mature scholars (without reference to age) whose ability to do productive research has been demonstrated.

[1] For the sake of completeness it should be stated that the "no change" method will also result in greater accuracy than the "judgment" method in any case in which the latter forecasts the general direction correctly but "overshoots the mark" by 100 per cent or more of the actual change.

The project for which aid is sought must be well under way and promise significant results. While the grants are awarded on a general competitive basis, preference is ordinarily given to projects involving two or more disciplines or likely to yield an important methodological contribution. Normally the maximum grant does not exceed $1,000. Wherever possible, institutions to which the applicants are attached are expected to contribute financially or with other special support. Grants may be used to help meet such items as the investigator's living expenses while in the field, the costs of travel, clerical or statistical assistance, photostating, printing and stationery; but ordinarily not for the purchase of books or manuscript materials. The next distribution of grants-in-aid will be made in March, 1931. Applications will be received up to February 1 of that year.

The Committee on Research Fellowships met on February 22, 1930, and considered 109 applications. From this number 30 awards, including one renewal, were made for 1930–31. The stipends attached to these awards aggregate approximately $80,000. These fellowships are designed primarily to promote the development of promising young research workers by giving them opportunities to carry on the investigation of some significant project in the field. While the term of the fellowships is normally one year, renewals for a second year are occasionally made. The stipend attached to each fellowship varies with the requirements of the Fellow. Cognizance is taken of the number of dependents, the amount of travel involved, and necessary clerical assistance and equipment. These fellowships are open to men and women, citizens of the United States or Canada, who hold the Ph.D. degree or its equivalent, and who are not over thirty-five years of age. While these requirements are not rigidly held to, exceptions are rarely made. The closing date for applications for 1931–32 is December 1, 1930. Awards will be made some time in February, 1931.

The third annual awards of fellowships in agricultural economics and rural sociology were made in March, 1930. From 83 applications fellowships were given to 24 graduate students. These Fellows will spend the next academic year pursuing graduate work at institutions scattered throughout the country. It is anticipated that partly as a result of the opportunity offered by these fellowships the growing demand for properly trained research workers in the fields of agricultural economics and rural sociology can be more adequately met. The stipends involved in the awards for 1930–31 aggregate approximately $34,000. Applications for 1931–32 will be received up to February 1, 1931.

The first annual awards of fellowships to southern graduate students in the social sciences were made on March 27, 1930. These fellowships were established through the generosity of the Julius Rosenwald Fund which made to the Social Science Research Council a grant of $50,000 to finance them. From 125 applicants distributed among practically all states in the South 27 students were given fellowships. They will be expected to spend the next academic year at some accredited graduate school working on social science problems of special interest to the South. Five of the new appointed Fellows are Negro students. Of the total number of Fellows appointed 10 were women and 17 men. The major purpose underlying these new fellowships is to attract a higher quality of scientifically trained personnel to work on important social problems facing the

southern section of the country. It is hoped that one of the important by-products of the fellowships will be the building up of a more enlightened social and economic leadership in the South. Applications for the coming year will be received up to February 1, 1931.

For further information regarding grants-in-aid, and for explicit directions and application forms for research fellowships, apply to Mr. Walter R. Sharp, Social Science Research Council, 230 Park Avenue, New York City. Information and application forms for the fellowships in agricultural economics and rural sociology may be secured from Dr. Edwin G. Nourse, Institute of Economics, 26 Jackson Place, Washington, D. C. Those interested in southern graduate fellowships should communicate with Mr. Will W. Alexander, 409 Palmer Building, Atlanta, Georgia.

MEETING ON BETTER DATA ON MERCHANDISE DISTRIBUTION

A dinner meeting of the American Statistical Association was held on Thursday evening, April 24, 1930, at the Aldine Club, 200 Fifth Avenue, New York City. One hundred and nine persons were present. Dr. Ernest A. Bradford, Director of the Bureau of Business Research of the College of the City of New York, acted as presiding officer.

The first speaker of the evening was Mr. Paul T. Cherington, Director of Research of the J. Walter Thompson Company. He spoke on the subject, "Distribution in a Buyer's Market." He began by pointing out that a great change has taken place in the degree of emphasis placed by manufacturers upon the selling side of their business. In the past, efficiency in *production* was stressed relatively much more than at present, the general theory being, that if goods could be produced, they could be sold. At present, many goods which are turned out by manufacturing concerns cannot be sold at a price sufficient to cover the cost, for they do not meet the style requirements of the consumers, and the number of incomes which, though small, leave a spendable surplus above bare subsistence has increased.

Mr. Cherington pointed out that available data do not permit one to measure, with accuracy, the potential demand of consumers in any given area. Income figures are decidedly incomplete, and such data as we have pertain primarily to individual incomes rather than family incomes. Purchasing, however, is done largely to meet the needs of families, rather than individuals.

The preliminary Census of Distribution has proved extremely helpful to students of marketing. A surprising fact revealed by this census is that a very large proportion of the total number of stores operating consists of units doing less than $5,000 worth of business per annum. Obviously, the profits on merchandise sold in an establishment of so small a size are not large enough to support a family. Such establishments must, therefore, be operated by persons having other sources of income.

Dr. Paul H. Nystrom, Professor of Marketing at Columbia University, the second speaker of the evening, agreed with Dr. Cherington in stating that more attention is now being given to the views of the consumers than was formerly the

case. For a hundred years after the appearance of Adam Smith's *The Wealth of Nations*, books on economics scarcely mentioned the consumer as playing a significant rôle. Even yet, in most economic texts, much more emphasis is laid upon production than upon consumption.

Dr. Nystrom next pointed out that the demand for consumers' goods is far from being static in its nature. Our institutions are ever-changing, and every change in economic life brings about a corresponding shift in demands for various classes of direct goods. Among such changes, Dr. Nystrom mentioned the following as being of great importance:

1. Urban population is growing faster than the rural population, hence the demand for goods suited to the needs of city dwellers is increasing in proportion to the demand for goods of the type used by inhabitants of rural regions.

2. The declining death rate among children has reduced the percentage of the population made up of children. The obvious result has been to cause the demand for certain types of goods used by children to shrink in comparison to the demand for similar goods used by adults.

3. As Dr. W. I. King in his recent book, *The National Income and Its Purchasing Power*, and Professor Paul H. Douglas in his *Real Wages in the United States, 1890–1926*, clearly show, the average rate of real wages has been increasing during recent years. This increase has brought forth a stronger demand for some luxuries and a weaker demand for necessities.

4. There appears to be a steady decline in the proportion of our inhabitants living in their own homes, and a corresponding increase in the proportion dwelling in apartments. This change has caused a shift in the type of furniture required and also in the type of apparatus utilized to assist in home work.

5. With increased prosperity, many home industries are declining. Even in the rural regions, people are coming to depend upon city bakers, and a larger and larger proportion of women's apparel is coming from the factories, while dressmaking, both at home and by dressmakers, is steadily declining.

6. The widespread use of the automobile is having a marked effect, not only upon home life, but also upon the mercantile industry. People now shop fifty miles away from home in place of five miles away from home.

7. College education has been expanded to cover a large section of the population and some of the results of this expansion are distinctly surprising. One of the most universal results is to emphasize the interest in style. Nowadays, clothing, furniture, and housing must all be in the latest fashion, if they are to be sold.

The Chairman supplemented Dr. Nystrom's remarks by calling attention to the fact that present estimates indicate that ten passengers now travel a mile in an automobile for every passenger that travels one mile on steam railways.

The third speaker of the evening was Dr. L. D. H. Weld, Research Director of H. K. McCann Company. His remarks were devoted in the main to demonstrating the marked success of the application of modern statistical methods to studies of market conditions. He showed that such procedures were now taking the place of the guessing processes which prevailed almost universally up to very recent times. He illustrated the crudity of methods previously used by pointing

out that it had been the custom frequently to use the sales of various articles as indicia of potential demand, without even reducing the sales to a per capita basis. He also pointed out the fallacy of using one general buying power index, such as magazine circulation, as a measure of the potential market for all commodities.

One test of potential demand in any area is the general buying power of that area. The only state in which the general buying power can be measured on the basis of actual sales is Pennsylvania, for which state figures are available showing retail sales by counties. The first step in attacking the sales quota problem scientifically was, then, to manufacture for Pennsylvania a composite index, variations of which correspond with variations in actual sales as reported by the Pennsylvania state authorities. The county index was worked up for Pennsylvania on the basis of a series of indicators, and this composite index, when compared with actual sales by counties, gave a correlation coefficient of +0.85.

The next step was to construct a similar composite index for the various states in the Union. When such an index was completed, it was compared with the sales of a group of manufacturers in these same states. These two variables gave a correlation coefficient of +0.92. This high degree of correlation seemed to indicate that the composite index worked out does actually constitute a satisfactory criterion of the buying power in the different states.

Dr. Weld showed how this general buying power index has been used to measure the quality of circulation of individual magazines. Some magazines follow buying power closely in that their circulations per one thousand inhabitants are high in states with high buying power, and low in states of low buying power. Other magazines have heavy circulations in states of low buying power.

In measuring the potential market for an individual commodity, it was pointed out that the correlation process may be used in determining the special factors that affect the sale of the commodity being studied. General buying power may be one of these factors. A study of the volume of sales of gasoline in different sections shows that this volume correlates closely with a composite index consisting of number of automobiles per thousand people, general buying power, the percentage of commercial vehicles, and the length of the driving season.

The fourth speaker of the evening was V. H. Pelz, Economist of the American Institute of Food Distribution. He called attention to the fact that there is still much statistical information which is desirable but not available. For example, we do not know how much cash-and-carry business there is. We have little information concerning the consuming habits of various classes of the public. The best existing information is the sample Census of Merchandising covering eleven cities. It is hoped that the census which is now being taken will fill in many of the gaps. We know in advance, however, that it will be weak in certain respects, for example, full-sized schedules are not demanded of either wholesale or retail establishments in towns under 10,000 inhabitants. Furthermore, no commodity groupings are given for the sales of stores selling less than $60,000 worth of goods per annum. The probabilities are, also, that the quality of the figures will be injured somewhat by the low wages paid to enumerators which amount to only 50 cents for a short schedule, or $1.00 for a long schedule.

At such prices, the enumerators cannot afford to put too much time on securing data for any one establishment.

The discussion was opened by I. G. Richardson of the J. C. Penney Company. He discussed the rates of growth of the various types of chain stores. He pointed out that the probabilities are that, of all the retail business transacted, chain stores had a larger proportion of the total volume in the larger cities than they did in the smaller cities and towns. Some classes of chain stores have been growing much more rapidly than others during the last decade. Those selling wearing apparel have expanded operations at an unusual rate. Chains of five and ten cent stores have also been growing very rapidly. Attention has been called to the fact that the sales, per store of individual five and ten cent companies, have been declining. The reason for this is not that the sales in any particular unit have been falling off, but that the chain business has been expanding to cover the small towns, thus taking in areas that will not permit of a large volume of business per store. The addition of these small stores is what has pulled down the average sales per store. Grocery chains have been growing at a very rapid rate. The tendency, at present, is to operate units in this field, combining the meat, grocery, and vegetable business. The slowest rate of growth in recent years has been in the chains of cigar stores.

The discussion was closed by Q. Forrest Walker, Economist of R. H. Macy & Company. He suggested the need for more information on the consumption habits of the people and the family budget. We need to know what the people do with their income. He stressed the fact that Dr. Nystrom has done more than anyone else to throw light on the facts in this connection.

The speaker also called attention to the very important fact that, even if the retailer finds out what the total sales of an article in a given area amount to, he is still unable to say what proportion of this total should represent his quota. Many people are competing for the consumer's dollar, and the individual merchant cannot tell in advance how many of the consumer's dollars he will be able to capture.

Mr. Walker also emphasized the point that style is growing more and more important and that it is now almost impossible to sell goods which are out of style. Formerly, when goods would not sell in New York, they were shipped to the smaller cities and sold well. At present, however, the smaller places are as up-to-the-minute on styles as are the people in New York, hence this transfer of goods is unprofitable.

WILLFORD I. KING, *Secretary*

MEETING ON CYCLICAL VARIATIONS IN EMPLOYMENT

The dinner meeting of the American Statistical Association was held on Friday evening, May 23, 1930, at the Aldine Club, 200 Fifth Avenue, New York City. One hundred and fourteen persons were present.

The presiding officer was Paul U. Kellogg, editor of *The Survey*. He called attention at the outset to the general tendency, during periods of unemployment, to investigate and map out programs which are forgotten with returning pros-

perity. As an instance, he cited the able report of the committee appointed by Mayor Mitchell of New York in 1914–15. This report was briefed and recirculated in the depression of 1920–21; but the recession of 1929–30 still found its recommendations largely unapplied. He expressed the hope that the various agencies set up during the last few months have more enduring quality; that we are not merely going through the old rigmarole, and that forces are at work which will lead to long and constructive planning.

The first speaker on the program was Meredith B. Givens, of the staff of the Social Science Research Council. He began by pointing out that, while statisticians are unanimous in the view that the volume of employment can be measured, they disagree vigorously on the possibility of measuring unemployment. Most of them feel that the latter process is very difficult, and for this reason they tend to confine their efforts to measuring changes in the volume of employment.

In determining changes in the volume of employment we have been compelled, in certain industries, to depend upon data collected for entirely different purposes. In the building industry, for example, we must estimate the volume of employment from the physical volume of construction. Current employment statistics are very scanty for most occupational groups outside of manufacturing, mining, and transportation. The American Statistical Association's Committee on Governmental Labor Statistics has been doing excellent work in urging the compilation of more inclusive statistics of employment, in securing comparability between figures from different sections of the United States, and in coupling state figures with those of the Federal Government.

Nevertheless, the measurement of changes in the volume of employment is no substitute for unemployment measurement. A sharp falling off of employment in one industry may be offset by an increase of employment in another industry, the result being that there is no net decrease of employment when both industries are considered as a unit. Even when employment is measured adequately in all fields of industrial activity there will be need for unemployment statistics in demonstrating the magnitude of the problem and in assessing the effectiveness of means of control and protection which may be developed. In the absence of adequate materials for unemployment measurement at the present time, it is likely that estimates and continual re-estimates are justified as instruments in pointing the way toward the improvement of our data.

A major difficulty in measuring unemployment arises from the fact that we have no accurate way of knowing how many people are available for employment. The 1930 Census will aid somewhat in estimating the number of persons who normally desire work. The basis for studies of employment and unemployment would be materially improved if it were feasible for the Bureau of the Census to work out a closer tie-up between the Census of Occupations and the Censuses of Manufacturing, Mining, and Agriculture. It is difficult at present to classify occupations according to industrial categories.

Recently we have heard much of technological unemployment. In studying this phenomenon, we should recognize that there are two distinct types:

(1) That which means a *shift in the kind of workers required in a given industry*, and

(2) That which means temporary or permanent *reduction in the number of workers required* in a given industry.

Unemployment measurement involves the second type since it is concerned with the number rather than the identity of the employed and unemployed. Statistical analysis of the impact of technological change upon employment rests upon the measurement of productivity changes. To this end it is important that we shall develop parallel series of statistics showing both production and labor time for an increasing number of processes and industries.

After the close of Mr. Givens' remarks, the presiding officer called attention to the unusual measures being taken during the present period of depression to understand and to remedy unemployment conditions. He stated that a study made by a civic committee in Lincoln, Nebraska, shows that, in that one small city, the annual loss due to wage earners being idle part of each year amounts to several hundred thousand dollars; perhaps a million.

He mentioned a number of undertakings in other localities, as Dayton, Ohio, Cincinnati, Philadelphia and New York State. A group of New England business men have set going a national study of seasonal fluctuations. The largest scale undertaking, however, is that of Swarthmore College, which has projected a scheme of research and expert service to run for at least a decade and to cover three great salients—stabilization, employment services and the protection of wage earners' families through employment reserves or insurance along original American lines.

The second speaker on the program was Paul H. Douglas, Professor of Economics at the University of Chicago and Acting Director of the Swarthmore Unemployment Study. He presented a series of charts regarding changes in the estimated volume of unemployment in each year since 1890. He pointed out, however, that the figures upon which his graphs are based did not include all fields of industry, as, in many lines, it is impossible to secure data of any value.

His method of procedure was to estimate the total normal labor supply for those industries from which information was obtainable. He next estimated the number of persons at work, and then assumed that the difference between these two quantities represented the number of persons unemployed.

Professor Douglas called attention to the fact that the percentage of the population which may be thought of as seeking gainful employment is not a constant, and does not necessarily follow a smooth trend. The reason that it does not is that, in periods of boom, it frequently occurs that many men and women are led to seek employment who, under normal circumstances, are not on the lookout for jobs. A curve representing normal labor supply, therefore, has humps in times of extreme activity of industry.

It is, in a way, paradoxical that the number of persons constituting the labor supply should increase in periods of extreme prosperity, for experience indicates that, in the country having high wages, the proportion of the population gainfully occupied tends to be smaller than where wages are low. The explanation of this apparently contradictory situation is as follows: During a boom, most people realize that the situation is not permanent and does not, therefore, represent a

definite increase in income levels. Workers, are, therefore, anxious to take advantage of high wages while it is possible to do so; and hence rush into industry with a full realization that their connection is likely to be temporary.

Professor Douglas found that peculiar conditions in certain industries made it necessary to compute unemployment by a method somewhat different from the standard. For instance, in coal mining, though most of the workers may nominally have jobs, they cannot work when the coal mines are not operating. In this field, therefore, unemployment was estimated on the basis of the time the mines operated.

The estimates made by Professor Douglas indicate that the average percentage of unemployment for manufacturing, transportation, building, and mining for the years 1896–1926 was 10.1 per cent, and that from six-tenths to two-thirds of all unemployment is due to secular and chronic causes combined, while the remainder is the result of the business cycle. He found no indication that there is any downward slope to the secular trend of unemployment.

He made the interesting suggestion that the fact that, in pre-war times, unemployment in the United States was nearly twice as great as in England might be due to the more changeable weather, which we have in this country, causing greater seasonal fluctuations in industry.

Professor Douglas expressed the opinion that the United States must follow the example of Europe and adopt unemployment insurance. In the later discussion, however, he stated that he thought it would be very inadvisable to follow the British and German method of making rates uniform to all employers in industry, and that, to be successful, the rate must be highest for the employer who had most unstable employment. In his opinion, a 4 per cent tax on wages would cover the cost in the United States for a 40 per cent coverage of the wages lost by the workers.

The third speaker of the evening was Ewan Clague, of the Institute of Human Relations, of Yale University. His address dealt with the study of the effect upon the employees of a large rubber company in New Haven of the closing down of the plant of that company. This company adopted a policy of paying to all long-service workers laid off, a termination wage, amounting, as a rule, to one week's wages for each year of employment with the company. The theory held by the management in paying this wage was that the employee who had been with the company for a long time had thereby, to a certain extent, lost the art of finding a job. Extensive inquiry among the workers receiving the termination wage seemed to indicate both that the money was rarely wasted and that its receipt did not in any way lessen the zeal of the workers in seeking a new position.

Those laid off by the rubber plant were largely Italians. The study was conducted one year after the closing of the rubber factory to see what had happened to the dismissed workers. The results of the investigation indicated that it is much harder for older workers to find a job than it is for younger workers. Only one-sixth of the older workers located new employment within two weeks, and 30 per cent had found nothing by the end of the year. On the other hand, all but 6 per cent of a younger group found employment by the close of the year. Forty-

five appears to be the dead line for factory workers in New Haven. Those above this age, when they did find a job, clung to it with tenacity. Younger men, on the other hand, were inclined to shift if they did not like the first position located. Women found work more easily than men. In general, the wages in the new jobs were not so good as the wages paid for the positions vacated at the rubber company. The few workers who did get better positions received but small gains in wages while those having losses, frequently lost heavily. In most instances, the new job was found either by going from plant to plant seeking work or through friends. Employment agencies gave but little aid. One man visited fifty-three different factories, some of them several times. Very few of those discharged left New Haven to seek work elsewhere.

About one-third of the families involved in the lay-off were home owners. Naturally, they found it very difficult to meet the payments on their homes while unemployed although the number of actual foreclosures was very small. Twenty-five per cent of the families had serious illness during the year.

One surprising feature revealed by the inquiry was the large amount of life insurance carried by these working-class families. Eighty per cent had some insurance. Most of the policies were of the "industrial" type. One family was found to be paying $8 per week for insurance. It was found that, whenever it proved impossible to pay all the insurance, the tendency was for the parents to drop their insurance, still carrying that on the children.

The discussion of the papers of the evening was initiated by Margaret H. Hogg, of the Department of Statistics of the Russell Sage Foundation, New York City. She expressed the view that the New Haven study reported by Mr. Clague has exceptional value because it covers a normal working class group, providing a contrast with studies of the ordinary unemployed who have a bias toward sub-normal efficiency. Miss Hogg plead for establishment of uniform procedure in unemployment surveys, and in particular that workers who are still usually employed should be carefully distinguished from those of middle-age and beyond who are no longer acceptable to employers, and who present a problem distinct from cyclical unemployment.

The formal part of the discussion of the papers was concluded by Willford I. King, of New York University. He endorsed the view of the chairman that the personal side of unemployment is extremely important, and also that it is just as important for industry to establish reserves for wage payments as for interest or dividend payments. His contention was that unemployment insurance, as applied in England and in Germany, has worked very badly indeed, cutting down production, and producing chronic depression in industry. He agreed with Professor Douglas that the situation would be greatly improved if employers were compelled to pay premiums in proportion to the amount of unemployment in their plants. He felt, however, that it would be far better if, instead of paying insurance, the employers, either voluntarily or under compulsion, gave continuous employment, for, in this way, production would be kept up and national prosperity would thereby be increased, since the buying power of the workers would not fall. He went so far as to propose that every employer be taxed upon percentage variations from the normal line of employment which he established well

in advance, and that this tax be made heavy enough to put out of business those employers who could not devise ways and means of stabilizing employment.

Dr. Royal Meeker, speaking from the floor, took issue with the last speaker and asserted that it was just as logical to insure against unemployment as to insure against accidents.

Mr. Charles W. Birtwell, of the Stable Money Association, then pointed out that, whenever unemployment insurance takes such a form as to lessen the desire of the worker for employment, it is incontrovertibly bad; but when its form is such as to stimulate the employer to keep his employees at work, it may be good.

Miss Mary B. Gilson, of the Industrial Relations Counselors, who has been carefully studying the British system of unemployment insurance, gave it as her view that it is not working out satisfactorily. It was her opinion that legislation should be enacted compelling the employer to keep at least part of his force steadily employed.

Professor Douglas, in replying to the suggestions regarding compulsory stabilization, contended that it is extremely difficult to stabilize any industry the products of which are subject to the whims of fashion, as, for example, is the case with production of women's clothing. He also raised the question as to what would be done in the case of a declining industry. He pointed out, furthermore, that unemployment insurance is needed for persons who lose their positions from causes entirely unrelated to cyclical and seasonal fluctuations.

Another speaker pointed out that, if a tax were laid upon those employers in the women's clothing industry who could not stabilize employment, the only result would be to increase the price of fashionable clothing, and that this increase in price would make fashionable clothing more exclusive, which would add to its desirableness.

Dr. Royal Meeker was of the opinion that it is futile to expect any extensive voluntary action in the direction of stabilization on the part of independent corporations. The only way to get action is through government compulsion.

Mr. Stanley B. Hunt concluded the discussion by calling attention to the fact that the Rayon manufacturers have been working in the direction of stabilization of employment, but have been hampered to some extent by the shifts in demand caused by changing fashions.

WILLFORD I. KING, *Secretary*

PROGRESS OF WORK IN THE CENSUS BUREAU

TASK OF TAKING THE FIFTEENTH CENSUS OF POPULATION

The great task of taking the Fifteenth Decennial Census of the United States is now practically completed, and the great task of tabulating and publishing the results is beginning. The 100,000 or more enumerators, with possibly a few exceptions, have completed the enumeration of their districts and turned in their portfolios to the local supervisors. And of the 575 local supervisors, all but about 90 at this writing (July 1) have reported that their work is completed and most of them have forwarded the boxes of schedules to Washington.

The difficulties of the task as well as the magnitude of it have been in some respects greater than ever before. There has been much complaint on the part of the enumerators about not finding people at home. It would seem that for a large proportion of the American population "home" is simply a place where they sleep at night. The typical apartment dwellers work during the day, get their meals out, and go to the movies in the evening; and the enumerator may call repeatedly without finding anyone in. This situation was by no means wholly unanticipated; and the Bureau made an attempt to provide for it by supplying individual slips and preliminary family schedules carrying the census questions, which the enumerator could leave when people were not in with the request that they be filled out and left to be called for, or in some cases mailed. But there was complaint that the people concerned paid no attention to such requests. In fact so far as the attitude of the people towards the census is concerned it seems as if the ignorance or suspicion of the foreign born and uneducated was not so great an obstacle as the indifference of certain classes of intelligent and well-to-do native Americans. However, it is still too early to generalize regarding matters of this kind. The Bureau will ultimately have in its records and correspondence a mass of valuable material in regard to the experiences, difficulties, and problems of census-taking. Especially valuable in this connection will be the reports which the supervisors are expected to submit describing the conditions and difficulties they encountered, indicating to what extent the various methods adopted proved successful or unsuccessful, and containing suggestions for the improvements of the methods for another census. Some of these reports have already come in and are found to be very interesting and instructive.

Some indication of the magnitude of the census and its complications may be obtained by considering the mass and variety of schedules or questionnaires which the Census Bureau has printed and distributed. Of the main population schedules, 3,206,000 copies were printed. As each schedule provides for 100 names, 50 on each side of the sheet, this was more than twice as many as would be actually filled out. But, of course, in this and other cases it was necessary to print a great many more than would actually be used, because every supervisor and every individual enumerator had to be supplied with a surplus in order to make sure that he would have enough. Of the individual slips, to which reference has been made, 8,765,600 were printed; of the preliminary family schedule, 4,000,000; of the absent family schedule (described below), 265,000; of the non-resident family schedule, 350,000; of the unemployment schedule (with spaces for 60 names), 700,000. Of the main farm schedule, 15,450,000 were printed, the surplus over the number actually required for enumerating 6,000,000 farms being used mainly for advance distribution to acquaint the farmer with the census questions. In addition the Bureau had to print a large number of illustrative examples, of test schedules, etc. And, of course, many varieties of schedules were used in the censuses of manufactures and distribution, which, however, as a rule were not handled by the regular population census enumerators.

The enumerator regularly carried in his portfolio twenty-three different kinds of printed forms including schedules, addressed envelopes, report cards, instructions, etc., not to mention blotters and a piece of twine.

ENUMERATING MIGRATORY FAMILIES

The absent family schedule, mentioned above, was used for the enumeration of families temporarily away from their usual place of residence at the time of the census, the idea being that this schedule would either be filled out by or for the Smiths or the Joneses prior to their departure if they were leaving shortly before the first of April or, if they had already gone and their temporary address was known, would be mailed to them to be filled out and returned by mail to the local supervisor. Theoretically this scheme should have caught all the absent families and credited them to the population of their respective places of abode. But practically, as we well knew, it would not do that. So as a complement to the absent family schedule the Bureau devised a non-resident family schedule, so-called. This was to be used where families were found temporarily residing in the locality but claiming permanent residence elsewhere. These schedules are sent directly to Washington, where they have to be allocated to the city or other locality which the family claims as its place of residence.

This allocation is going on here in the Bureau at the present time, and, as was anticipated, is something of a job, since these non-residents if possible must be allocated not only to the city but to that precise district of the city in which their residence is located; and to do that readily one must know something more than the number and street because with that information only it is difficult to find in what block the given number is located. So the form used required them to give not only the number and street on which they lived, but the streets between which they lived, as, for instance, 937 Main Street, between Spruce and Pine Streets. But, of course, in very many instances they failed to supply that information.

It may seem that these are minor details of census-taking, but they illustrate some of the difficulties of enumerating a very mobile population in their "usual places of abode."

POPULATION ANNOUNCEMENTS

It is too early yet to judge of the quality of the Fifteenth Census. Here in the Bureau we believe that we had on the whole as good a class of supervisors and as good enumerators as any census has had, if not better. And until we have unpleasant, convincing evidence to the contrary we shall believe that we have taken as good a census as was ever taken, if not better. But no doubt there will be plenty of room for improvement another time.

In one respect at least this census has already made a record, and that is in the promptness of the population announcements. At this writing (July 1) the population of 93 cities having a population of over 100,000 each has been announced. At the corresponding date following the census of 1920 the announcements covered only 10 cities of this class. At the same time the population has been announced for a large proportion of the smaller cities and of the counties. In all, the announcements thus far made cover probably over 80 per cent of the total population of the United States.

One reason for this improvement is found in the practice, introduced at this census for the first time, of having the local supervisors count the names and make the announcements as soon as the canvass is completed and before the schedules

are sent to Washington. Another advantage of this arrangement is that it gives an opportunity for charges of a deficient count, or local "kicks" of any kind, to be investigated at once while the schedules are still there and before the enumerators have been paid off and discharged. Five days are allowed for complaints, and then if none worthy of consideration materialize the schedules are forwarded to Washington. Of course the count by the supervisors is checked in Washington, and the population announcements thus far made are provisional and subject to revision. But rarely will there be any material change in the figures.

The total population of the United States will be known by the time this article is in print.

UNEMPLOYMENT ANNOUNCEMENTS

In addition to the population announcements the Bureau has been giving out some statistics from the census of unemployment, limited for the time being to the number of persons reported on the unemployment schedule as without a job, able to work and looking for a job. These might be termed the jobless. Other persons were reported as having a job but as laid off or not working for some reason. Whether some of these, and if so which of them, should be included in the total unemployed is a question on which opinions may well differ. There is, so far as the writer is aware, no generally accepted and exact definition of the term "unemployed." It is safe to say that no definition ever formulated would include as unemployed all the persons reported on the unemployment schedule. On the other hand, it is probable that almost any reasonable definition would include some of those reported as having jobs but temporarily laid off. Ultimately the Bureau, according to present plans, will include in the final report all persons reported on the unemployment schedules classified in detail under various heads but perhaps without undertaking to designate the classes that should be included in the total unemployed. The figures for the jobless were given out first because they seemed most significant and could be easily arrived at by a hand count of names on the schedule made by the supervisor.

For lack of time and space any report of progress on the censuses of agriculture, manufactures and distribution will have to be postponed to a later issue of this JOURNAL.

J. A. H.

MISCELLANEOUS NOTES

The Austin (Texas) Chapter.—The final meeting of the Austin Chapter for the academic year 1929–30 was held on May 14 with twenty members and guests present for luncheon. The feature of the meeting was an address by Mr. C. A. Duval, of the Department of Economics of the University of Texas, on the subject, "Possibilities of Gains and Losses Due to the Stabilization of Farm Prices."

Mr. Duval stated that studies by Dr. Ezekiel and Dr. Black tend to show that the stabilization of corn and oats prices around their normal trend of prices will result in gains in gross income to the farmer; that the stabilization of the price of potatoes is too expensive, on account of the cost of storing; and that the stabilization of cotton and wheat prices, because of the concave shape of their demand curves (supply-price relationship), will result in a reduced gross income. But Mr. Duval's own study tends

to show that the conclusions drawn by Dr. Ezekiel and Dr. Black are misleading in regard to wheat and cotton. Both the gross incomes and the net incomes to the producers of cotton and wheat will be increased as a result of the stabilization of their prices, thinks Mr. Duval.

For cotton and wheat, Mr. Duval's study shows the following:

1. The demand schedules will be much more like the straight-line schedules of corn and oats if the speculative holding is reduced or eliminated.

2. The stabilized normal price will be 5 to 10 per cent higher than the normal price trend. The present low trend of prices is due to an excessive amount of acreage in cotton and wheat; this excessive acreage is due largely to the fluctuating prices of these products. A price six cents above the normal trend price for cotton causes an acreage increase of about 12 per cent the following year, if the acreage the first year was normal; whereas, a six-cent decrease will cause only an 8 to 10 per cent reduction in acreage.

3. The average cost of producing the present crop is greater than the cost of producing the same size crop under stabilization. This is due to the fact that acreage added above the normal trend is usually of a poorer grade than that acreage abandoned below the normal acreage trend.

4. Production costs and marketing costs will probably be reduced under stabilized prices.

5. The stabilized normal price will be higher than the fluctuating normal price; and the cost of producing this smaller amount would be correspondingly less.

Following a discussion of Mr. Duval's paper, officers of the Chapter for the ensuing year were elected, as follows: President, Dr. F. A. Buechel, Statistician and Assistant Director, Bureau of Business Research, University of Texas; Vice-President, Mr. F. E. Finley, of the Texas Field Staff of the United States Bureau of Agricultural Economics; Secretary (reëlected), Mr. C. D. Simmons, Professor of Business Statistics, University of Texas.

The Boston Chapter.—The Third Annual Meeting of the Boston Chapter of the American Statistical Association was held at the Boston City Club on Monday evening, April 14, 1930, and was attended by 42 members and guests. Dinner was served at 6.30 P.M.

The business session was called to order by the President, Mr. Charles F. Gettemy, at 7.45 P.M. The following were elected to serve as officers of the Chapter for the ensuing year: President, Professor Davis R. Dewey, Massachusetts Institute of Technology; Vice-President, Dr. John D. Black, Harvard University; Secretary-Treasurer, Roswell F. Phelps, Director of Statistics, Massachusetts Department of Labor and Industries; Counsellors, Leroy D. Peavey, Babson's Statistical Organization, Wellesley Hills, Massachusetts; Edward A. Filene, William Filene's Sons, Boston.

The meeting was addressed by Dr. O. M. W. Sprague, Professor of Banking, Graduate School of Business Administration, Harvard University, on the topic, "International Financial Coöperation." The following summary of Dr. Sprague's address was prepared by Professor Gilbert H. Tapley:

Dr. Sprague contrasted international financial problems with the regional financial problems of the United States, and explained briefly how the transfer of gold from one country to another had aided in maintaining international economic equilibrium. He mentioned Europe's complaint that, by sterilizing much of the gold which had been sent to the United States, we had greatly increased the European financial problem.

In discussing the effects of the Young Plan and the Bank of International Settlements, he made it clear that the reduction and definite fixing of reparations payments were much more important than any changes made in the mechanism of payment. Though settlements are to be made through the new bank, there will be no radical change. The Bank of International Settlement will be an accounting agency rather than the one to decide when exchange should be purchased.

The new bank will bring out the German Bonds to provide cash in the near future, but will probably not have a large voice in major decisions, as the banks in the various countries will probably decide the time of issue and the rates to be used.

It will, however, serve as a forum in which agreements will be made by the representations of the various countries. This will be new only in that it provides permanent machinery for such conferences and avoids the necessity of creating new arrangements for each issue.

Other activities of the new bank will probably be more important, for it will provide a meeting place for the heads of the larger banks of issue. While there have been many such meetings in recent years, they were largely unknown before the War, and they are more likely to continue with this new organization. It will bring these men together in person, avoiding the necessity of formal correspondence on major problems, and, though no particular action of the bank may result, the bank should be credited with the benefits arising from such personal contact.

Beyond the field of reparations, it is difficult to see large activities for the bank. If gold balances were maintained by the nations, settlements could be made by book entries, thus avoiding the cost of shipping gold, but the expenditures for shipping gold are not now large.

It could help in preventing unbalanced situations and aid in restoring balance if the granting of small credits were all that is necessary, but frequently neither a small nor a large credit by some outside source would be adequate. The over-extended condition of a country or the over-supply of the nation's principal product usually requires liquidation rather than more credit, though the nation, like a business in similar circumstances, thinks that it needs only more credit.

If it has large funds at its disposal, the bank will receive many such requests, and it will be difficult to decide whether it is desirable to grant credit, also difficult to refuse when proper to do so. Fortunately, the bank will not have large funds at first, perhaps 200 millions, and no means of acquiring large funds for a considerable period.

It can, however, perform a large service of a negative sort. Before the War there was a tendency for some banks of issue to maintain large stocks of gold. A scramble for gold is now clearly undesirable, for the future output is uncertain and should be conserved. Reserves should be reduced and this cannot be done unless all or nearly all countries will permit free gold markets. The International Bank should provide an agency for spreading such fundamental financial principles.

Before the War, the London money market could always finance desirable foreign loans. Local needs never absorbed all the available credit. After the War, the United States made large foreign loans for a while, then stopped, as domestic needs expanded, using all our credit and even withdrawing funds from abroad.

The United States is not yet, therefore, a reliable source for foreign financial needs, but this is a matter outside the province of the International Bank. No single agency can correct a situation in which credit at 8–10 per cent is absorbed by a domestic security market.

In sum, while the bank may be expected to afford assistance in cases of minor difficulty in the weaker debtor countries, its most valuable service will perhaps be

found in the promotion of the effective use of monetary gold stock throughout the world.

The address was followed by discussion, in which representatives from two Boston banks participated. The meeting adjourned at 10.15 P.M.

The Chicago Chapter.—A dinner meeting was held on April 30. After a short business session consisting principally of announcements regarding the annual meeting to be held in May, the Chairman introduced Dr. Edith Abbott of the University of Chicago, who spoke on problems in social work and social statistics. Dr. Abbott gave the group a description of social administration, discussing the differences between workers in that field and social science workers. She pointed out that the problems of social administration were practical, everyday problems dealing with the lives of specific human beings. From this view of the problem, she illustrated by means of statistical studies prepared for social science purposes how inadequate such studies are when applied in any specific case of social administration.

Dr. Abbott discussed quite thoroughly the work which has been done and is now under way in the preparation of statistics which may be valuable in the solution of social administration problems. She feels that very little has been done toward the collection of statistics which are useful in social administration work. She discussed, briefly, the work of the United States Children's Bureau which has been under her direction, leaving the details of the statistics collected and their uses to Mr. A. W. McMillen, who continued the discussion for the remainder of the meeting, illustrating with charts the problems encountered in social administration. He also showed what statistics had been collected and, while they have not been applied as yet to practical problems, he felt that they were making some headway in developing data for social administration purposes.

The discussion which followed these talks consisted principally of matters relating to the specific work of social administration rather than to the statistics themselves.

The annual meeting of the Chicago Chapter was held on May 21 and was attended by about forty-five members.

After a report of the Chapter's financial condition by the Treasurer, the election of officers for the 1930–31 season was held with the following results: President, Mr. H. G. Pett, Manager, Department of Research and Statistics, Federal Reserve Bank of Chicago; Vice-President, Mr. D. R. Cowan, Commercial Research Department, Swift and Company; Secretary-Treasurer, Mr. H. B. Stair, Assistant Statistician, Illinois Bell Telephone Company; Board of Directors, Mr. E. J. Fowler, Statistician, Commonwealth Edison Company; Mr. N. C. Murray, Statistician, Clement, Curtis and Company; Mr. E. M. Keays, President, E. M. Keays Company.

Speakers of the evening were Mr. H. C. Campbell, Director of Research, J. Walter Thompson Company, Chicago; who spoke on "Advertising Agency Research," and Mr. E. M. Keays, President, E. M. Keays Company, Milwaukee, whose subject was "New Kinks in Business Cycles as Related to Investment Analysis."

Mr. Campbell first discussed briefly the factors involved in advertising: the advertiser, the owner of the media through which the public is reached, and the advertising agency or connecting link between the advertiser and the media. Next he sketched the typical organization and functions of an agency. The department in charge of the contact function is the business creating organization of an agency. The media department buys space while the research department gathers and analyzes information as an aid to the work of the agency as a whole.

The Research Department's functions are: trade and consumer investigation, merchandising analysis, and miscellaneous research. Trade and consumer investigation is carried on privately to ascertain copy appeal and effectiveness and determine when to advertise. Merchandising analysis is necessary since many clients, particularly small ones, do not analyze their own merchandising problems but call on the agency for advice. Miscellaneous research includes the provision of historically correct material for copy and analysis of the characteristics of readers of various publications. Mr. Campbell stated that simple statistical methods were used and very few mathematical methods applied in advertising agency research work. As in all research work, the biggest problem is the determination of significant factors.

Mr. Keays reviewed the discussion of the New York Chapter, "Is There a Business Cycle and If So What Is It?" and then outlined some of the factors involved in the business cycle. According to Mr. Keays, three elements are involved in each cycle: material development, a state of mind, and an accident.

The Cleveland Chapter.—The annual meeting of the local Cleveland Chapter took place on May 12. President H. B. Flinkers presided. There was a good attendance, and the program was as follows: "The 1930 Census in Cleveland," Mr. H. W. Green; "Population Estimates," Mr. D. A. Hill; "On Manufactures and Distribution," Mr. G. W. Schreech; "Sidelights on the Steel Industry," Mr. John W. Love.

At this meeting the following officers were elected for next year: President, H. Wembridge, The Joseph & Feiss Company; Vice-President, D. A. Hill, Ohio Public Service Company; Secretary, D. C. Elliott, Midland Bank.

The Business Statistics group decided to hold regular luncheon meetings during the summer if possible, and accordingly the first summer meeting was held on June 24.

Activities in the Detroit Area.—Members of the Association from Detroit and vicinity held their third meeting of the year May 8, with 34 members and guests present. The Chairman for the evening was Mr. Carl F. Behrens, Economist for the Union-Guardian group.

The general topic for discussion was forecasting for a particular business. Mr. Roland Hill, Statistician for the Dairy Products Research Bureau, described the methods used in forecasting the sales of fluid milk in the Detroit area. Sales were forecasted for individual companies and for groups of companies. It was found that the projection of a "normal" based upon secular trend and seasonal variations would give monthly forecasts of sales with an error of about 4 per cent. The divergence of actual milk sales from this projected "normal" was accounted for almost entirely by employment and business conditions in Detroit. When the projected normals were adjusted for predicted business conditions, based upon automobile production, errors in the monthly forecasts were reduced very considerably.

The second speaker, Mr. C. C. Ashbaugh, Statistician for the Michigan Bell Telephone Company, described the methods used by that organization in setting up their yearly budgets. He also described the company's index of Detroit business activity, which, at the end of March, was approximately 6 per cent below normal.

Los Angeles Chapter.—A series of meetings devoted to the discussion of various economic problems affecting Los Angeles and Southern California have been inaugurated by the Los Angeles Chapter. The monthly meetings during the spring quarter covered, respectively, (1) Los Angeles' aspiration to become a call money center, (2) building-loan operations in California, and (3) chain grocery operations in South-

ern California. All three meetings were dinner meetings held at the University Club of Los Angeles.

The speaker at the March 19 meeting was Mr. George H. Ellis, of the Bank of America of California, who presented a talk on "Los Angeles as a Prospective Call Money Center." Mr. Ellis described the organization and functioning of the New York Call Market, and told something of the progress and difficulties of such markets in Chicago, Detroit, San Francisco, and other cities. He then pointed out the problems which would confront a call market here, such as the varying amounts of money needed, the distance from the major money market in New York, and the problem of handling collateral, and described a plan which had been discussed at one time by Los Angeles bankers. Discussion of the problem was led by Mr. F. E. Sanford, Secretary of the Los Angeles Stock Exchange, and Mr. J. R. Douglas, Vice-President of the Security-First National Bank of Los Angeles.

On April 23 the Chapter devoted its attention to building-loan problems. Mr. W. F. McDowell, Past President of the United States Building-Loan League, and now Vice-President of the Pacific Coast Building-Loan Association, addressed the meeting on the topic of Building-Loan Associations as Business Stabilizers, pointing out that the building-loan plan leads to home ownership and systematic saving, providing a reserve against which the average man can draw in hard times, and turning him to saving rather than careless spending when his income is satisfactory. Mr. McDowell's talk was followed by a more technical explanation of the various types of associations operating in California, by Mr. A. A. Anderson, Secretary of the Pacific Coast Building-Loan Association. He classified them as the mutual type, the guarantee capital (non-mutual) type, and the mutual-guarantee type, and proceeded to analyze their respective business methods, profits, stability, and investment desirability. A general discussion of the functions of savings banks, insurance companies, mortgage banking institutions, and private lenders followed.

The May 21 meeting was given over to a discussion of the problems of chain-store control. The speaker was Mr. Edward Dale, Vice-President of Safeway Stores, Inc. Mr. Dale traced the growth of the chain store idea, spoke of the opposition of many independent dealers to the chain store, and illustrated by numerous anecdotes the problems met in the development of the Safeway chain. Statistical evidence was introduced to demonstrate that the chain store succeeds only by continually improving its service, cutting its costs, and devising new appeals to the public. Numerous comments and questions by those present rounded out a highly informative meeting.

These three meetings were attended by from fifteen to twenty members each. The Chapter is glad to report a slow but steady increase in membership during the past year. The regular monthly meetings have been adjourned for the summer.

Activities in the Pittsburgh Area.—The Association membership in Pittsburgh, as has been pointed out before, has been working in close partnership with the Research and Statistics and Manufacturers Committees of the Pittsburgh Chamber of Commerce. As such the membership has been very much interested in the subject of the stabilization of unemployment. We have been fortunate in bringing to this city outstanding speakers who have had experience in other parts of the country with this problem. Dr. Francis Bird, Professor of Commerce at the University of Cincinnati, detailed the methods by which the city of Cincinnati has been able to eliminate much local unemployment and stabilize relief expenditures for families temporarily destitute. Mr. Morris E. Leeds, President of Northrop & Leeds of Philadelphia, discussed the approach which the city of Philadelphia, together with the University of

Pennsylvania, is making for a solution to this problem in that city. Finally, Professor William M. Leiserson, of Antioch College, Ohio, gave a very illuminating talk on the economic problems to be solved in any analysis and solution of this general problem of unemployment.

All of these speakers brought out the fact that a more active interest in this matter on the part of the public and industry is essential if there is to be any real progress toward evening out the cycles of unemployment. It was felt that the most practical way of getting at this problem is first of all to set in motion machinery capable of collecting facts and interpreting them. Secondly, to obtain the coöperation of the industrial executives of the community so that they would not only appreciate the situation but could themselves set in motion any activities that would tend to alleviate the trouble. Attention also was given to the problem of vocational training so that there would not, in future, be more men prepared for one line of activity than could reasonably be expected to find employment in that field. Also it was felt that greater effort should be made to train the future members of the laboring group to adapt themselves more readily to different lines of endeavor so that should they be unnecessary in one business they could be profitably employed elsewhere. A final phase of the matter involved plans whereby the present group of skilled and unskilled labor could be assured of a reasonably continuous income by means of coöperative insurance schemes. This latter problem is of particular importance as it deals with the immediate situation and provides not only a steady income for the mass of people but likewise insures the producers a more continuous source of purchasing power.

United States Bureau of Labor Statistics.—The results of the survey of the standard of living of employees of the Ford Motor Company in Detroit were published in the June issue of the *Labor Review*. This survey was made at the request of the International Labour Office as part of a study of the living standards of $7 a day men in Detroit, in comparison with the costs of an equivalent standard of living in various European cities. The European part of this work is now being done by the International Labour Office.

The Bureau is now engaged in the preparation of a report dealing with the activities of governmental labor agencies. During the past several years the scope of the work of the various federal and state agencies dealing with labor matters has constantly expanded. The contemplated report will seek to give a brief non-critical account of what these agencies are doing.

The report on the recent investigation of employment, wages, prices, living conditions, etc., in the Hawaiian Islands, made by the Bureau, is being prepared and will be sent to press shortly.

A field survey of labor productivity in the sheet and tin-plate industries and the semiannual cost-of-living study are in progress.

Field work was continued during the past quarter on the surveys of wages and hours in the cotton manufacturing, cotton dyeing and finishing, woolen and worsted manufacturing, and rayon industries, and work was begun on similar surveys for the lumber, hosiery and underwear, and sugar-refining industries and for the organized trades.

The gathering of data on wages and hours in cigarette making has been completed and the data are being tabulated.

During the quarter under review, the labor turnover data compiled by the Bureau for separate industries were augmented by the addition of rates for the furniture industry. The companies reporting to the Bureau on labor turnover now number over 3,000.

The employment figures published by the Bureau now cover crude petroleum production in addition to manufacturing, mining, quarrying, public utilities, trade, hotels, and canning and preserving. Indexes for each of these industries except crude petroleum production are now being published. The Bureau attempts to secure representation of at least 50 per cent of the total number of employees in an industry and in some industries as high as 80 per cent of the total number of employees are covered.

United States Women's Bureau.—Three studies have been completed recently by the Women's Bureau of the United States Department of Labor. In a report on *Industrial Accidents*, an analysis of the state accident reports brings out the fact that only 21 states published accident figures by sex at any time during the period 1920 to 1927 and that such reports were, in the majority of cases, inadequate, unstandardized, and non-comparable.

In spite of these difficulties, the published data on accidents to men and women in 21 states in 1927, or the most recent previous year since 1920, do give certain important facts on accidents to women, and these are analyzed in the report. Among them it is shown that although accidents to women were actually and relatively fewer than those to men, and although women constituted a smaller proportion of the total number injured than of the total number gainfully employed, the numbers of women injured were large. Machinery was the chief cause of the accidents to women in three of the five states reporting, and falls of persons and handling of objects also were frequent causes. For men the handling of objects and, though generally to a much less extent, falls usually caused more accidents than did machinery.

The second survey dealing with *Women Workers in Laundries* covers the hours, wages, working conditions, and the composition of the working force, including such facts as nativity, age, marital condition, and length of service of the women. The laundry industry is of particular importance to the woman worker, as two-thirds of its operatives are women. Recent years have witnessed a rapid growth in its size and efficiency and there is an increasing tendency toward consolidation.

The Women's Bureau study is based on records obtained from 290 laundries in 23 cities situated in 17 states. Only the larger and more representative cities were surveyed, and the laundries selected were those in which a considerable number of women were known to be employed. The total of employees in the laundries was 24,337, of whom 19,758 (81.2 per cent) were women. Although only 23 cities were visited, the number of wage earners included is more than one-seventh of the number reported for the entire industry in 1925, and all laundries in these cities could not be scheduled.

The particular hazards of the laundry industry are too great temperature, unguarded machinery, and wet and slippery floors. Nearly one-half of the plants visited were reported by the investigators as being "comfortable" and about the same proportion as being "warm" or "hot." Seven laundries of the 290 had flat-work ironers without guards and 92 had extractors without guards. In 44 laundries no guards or presses whatever were found. In spite of hazardous or uncomfortable conditions, the labor turnover was low. Over one-half of 1,296 white women interviewed had been in the industry off and on for 5 years and more, and 15.4 per cent of the women had worked off and on for 15 years or more.

Wages and hours in the laundry industry present no startling deviation from the average. The median of a week's earnings for white women was $16.10. In the East 58.9 per cent of the women had working hours of over 8 and under 9; in the Middle West, 67.9 per cent worked a 9-hour day. In the West 97.2 per cent worked 8 hours a day, and 82 per cent in the South had a working day of over 9 and including 10 hours.

A report on *Industrial Homework* discusses the extent, character, and causes of homework, the inadequacy of its earnings, the difficulties of its regulation, and the hazards to public health. The most important parts of the report and recommendations of the Committee on Industrial Home Work of the Association of Governmental Labor Officials of the United States and Canada are given verbatim, and a list of reading references has been compiled.

Children's Bureau.—As part of the program for the development of current statistics in child welfare and related fields, plans have been completed for the Children's Bureau to take over on July 1, 1930, the program of registration of social statistics, which has been carried on for the last two years as a joint project of the Local Community Research Committee of the University of Chicago, and the Association of Community Chests and Councils. The need of uniform social statistics as a basis for social planning, not only for a particular locality but for the state and the Nation, makes this undertaking, which marks the beginning of such statistics, of the greatest importance.

Recommendations on Unemployment Statistics.—In the hope of winning the coöperation of governors in the extension of the monthly collection of employment statistics, the recommendations of the Committee on Governmental Labor Statistics of the American Statistical Association were submitted to the chief executive of every state in the week before the National Conference of Governors convened in Salt Lake City on June 30. Governor Roosevelt gave special emphasis, in his address at the Conference, to the necessity for better statistics of employment, and it is evident that in subsequent discussion considerable attention was given to the need for strengthening state provision for labor statistics.

At the annual meeting of the Association of Governmental Officials in Industry of the United States and Canada, held in Louisville in May, a resolution was adopted, endorsing the principle of collection and publication of payroll figures by industries, urging coöperation with the United States Bureau of Labor Statistics and recommending that procedure be made uniform between the different states in order to secure comparability of figures. This is essentially the plan which is being advocated by the Committee on Governmental Labor Statistics, and members of the Committee were active in support of this action by the Association. A tentative draft of the recommendations to be made to governors was informally circulated among members of the Association at their meetings, in order that they might be informed of the Committee's action as it might affect their work, in their own states, and that they might have opportunity to suggest changes in the statement. Their endorsement of the principles of the plan was virtually an acceptance of it, and a number of these officials have coöperated with us in an effort to interest their governors.

The Brookings Institution.—At the June dinner of the Brookings Institution it was announced that the following research fellows had been appointed for the coming academic year: G. Lyle Belsley, reappointment; Julian S. Duncan, of the faculty of Bryn Mawr College; Raymond W. Goldschmidt, Institut für Finanzwesen, Berlin; A. C. C. Hill, Jr., London School of Economics; Susumu Kobe, University of Michigan; Yau Sing Leong, reappointment; Martin L. Lindahl, of the faculty of the University of Michigan; Meredith V. McDougal, Johns Hopkins University; Nathan Otton, the Statistisches Reichsamt, Berlin; Charles Norman, Leland Stanford University; Eleanor Poland, of the research assistant staff, Harvard University; Barkev S. Sanders, of the Children's Bureau; Carroll K. Shaw, University of Illinois;

Carel Jan Smit, University of Amsterdam; Henry J. Wadleigh, Oxford University; Marion L. Wadleigh, University of New Zealand; Kenneth O. Warner, reappointment.

H. G. Moulton is spending the summer and fall in the Orient in connection with a study of the economic position of Japan.

Of the Institute of Economics staff, L. R. Edminster taught during the summer quarter at the University of Virginia; Thomas Walker Page taught at the University of California; and L. S. Lyon is giving a course in the National Institute for Commercial and Trade Organization Executives, at Northwestern University. Leo Pasvolsky has spent the spring and summer in Europe studying the problem of currency stabilization; and C. O. Hardy is spending the late summer and fall in Europe, in connection with a study of central banks. Sumner H. Slichter, who has been with the Institute on a year's leave of absence from Cornell University, has accepted a position with the Harvard Graduate School of Business Administration. Marie Butler has joined the Institute staff as editor.

Of the Institute for Government Research staff, W. F. Willoughby was in Europe during the spring, and Fred M. Powell during the summer. Paul B. Betters, formerly on the faculty of Syracuse University, has accepted a position with the Institute.

Guide to Social Welfare Statistics.—The Welfare Council of New York City has recently published a 416-page *Guide to the Statistics of Social Welfare in New York City*, which is presumed to index all the principal existing items of statistical importance concerning the welfare of the people of the city. It is modeled on the *Guide to the Current Official Statistics of the United Kingdom*. Miss Florence DuBois of the research staff of the Council is the compiler.

PERSONAL NOTES

James Arthur Harris

Dr. J. Arthur Harris, Professor and Head of the Department of Botany in the University of Minnesota, passed away on April 24, 1930, after a very brief illness.

Born in Plantsville, Ohio, on September 29, 1880, he early moved with his family to Kansas where, during his boyhood, he was associated with pioneer conditions that had a very profound influence on the rest of his life, for he became a pioneer of indomitable spirit in scientific work until the day of his untimely passing.

His studies as an undergraduate were pursued under the care of the faculties of Kansas University. At the age of twenty-one years, he had taken his A.B. degree and had to his credit four worthy publications. The latter are stimulating in the glimpses they give of his very keen power of observation and rational deduction at that early age concerning natural phenomena. His love for an understanding of natural laws ran sufficiently deep to make every observation of great value, and he quickly grasped the importance of the description of variation in quantitative terms. It became his chosen life task within the next few years to attempt to help place Biology alongside of Physics and Chemistry in the ranks of the so-designated "exact" sciences. He pursued this course with untiring energy, and against very heavy opposition in the early days when many another lacking his conviction, his courage or his industry would have given up the ideal in disgust.

His first professional appointment was that of Assistant Botanist at the Missouri Botanical Gardens from 1901–03. During that period he completed the requirements for his A.M. degree at Kansas University (1902), and in 1903 received the degree of Doctor of Philosophy at Washington University. In 1904 he accepted the responsibilities of Librarian at the Missouri Botanical Garden which position he held until 1907, when he became Botanical Investigator to the Station for Experimental Evolu-

tion of the Carnegie Institution at Cold Spring Harbor. From 1903–07 he also combined an Instructorship in General Biology at Washington University with his other work.

As the opportunities and facilities for the pursuance of his research broadened, so did his tremendous energy intensify and his scientific writings increase in number and power. Paying computing assistants to the limit of all he could spare from his own salary in addition to those available to him from his research budget, he himself spent long and weary hours computing on his Brunsviga. Finding himself physically capable of working thirteen hours a day, nothing could excuse him from that period of devotion *per diem* to his labors.

A lover of the quantitative and precision of description in biological work, he nurtured a profound admiration for Karl Pearson, whom he regarded as "the Giant of our time in Biological Science." In 1908–09 he availed himself of an opportunity to carry his studies to within the walls of the Galton Laboratory. Those days enriched his enthusiasm for his work in a very high degree. He retained till his death his veneration for the mind that had so inspired him in his biometric work, and passed that feeling on to his students.

In 1921 his was the honor to receive the Weldon Medal and Weldon Memorial Prize from the University of Oxford.

In 1924 he was called to the post he held at the time of his death. His coming to the University of Minnesota proved to be one of the richest gains in the history of that institution. The progress of the development in teaching and inspired diligence in research in the Department of Botany remains a tribute to his keeness of judgment, administrative power and quality of leadership. With rare rapidity he won the love and admiration of his colleagues and students. All acknowledged him master. Students sat "at his feet" eager to absorb the learning and the enthusiasm that so freely flowed from his most earnest teaching. His public lectures won him esteem from all for the richness of his philosophy, and the subtlety of his humor. In University committees he was ever regarded as a most valuable adviser, whether the subject involved administration, teaching, research or judgment of men.

Although professionally a Botanist, the whole field of Biology was his realm for the application of the statistical method. Extensive series of data that became available and presented opportunity for new achievements in biological science always claimed his interest. But above all, he himself was a compiler of original records. Whenever he was free from the office or laboratory, he could be found in the field with a party of young men of high calibre assisting him in the routine of observation, collection and determination.

A student of Ecology, his greatest interest always centered in the plant or organism living in its natural surroundings. His studies of the chemical and physical properties of plant tissue fluids in relation to the plant's environment had taken him for the last nine summers into the deserts and alkali lands of Arizona and Utah. He reveled in those expeditions with the true spirit of the pioneer and the lover of nature. Previous studies had oftentimes led him to struggle through the morasses of the Dismal Swamp and the Everglades, through the montane rain-forests of Hawaii, and through the forests and deserts of Jamaica. Before his death he entertained high hopes of carrying his physico-chemical studies of plants to the torrid regions of the Dead Sea. For the many expenditures which he bore from his own salary for the financing of these expeditions, he claimed but one reward—the advancement of science. His joy lay in the searching out of those truths of the universe about us which might be put down for the benefit of future generations.

As a biometrician he was a truly inspired man. Without any college training in mathematics to aid him, he intuitively grasped the philosophy of mathematical thought, astonishing advanced students of that science with his power to grasp and logically employ an end equation, the derivation of which he would have to leave to others. With practically nought but the symbolism of algebra to express his mathematical conceptions, his contributions to the theory of statistics redound to his praise. And surely one of his great contributions to biology has been his translation of complex mathematical equations into simpler forms that were more readily comprehensible to workers in that great field.

To him the gift of great intellectual power meant but a greater obligation in service. His own thirty years of endeavor in science are only in part recorded in his three hundred papers.[1] One may not measure as yet the full influence or significance of his works, for we are too near in time and yet too far behind him in understanding. The full loss to science in his sudden passing may never be estimated, for he was still advancing toward a prime yet many years distant.

An excellent host, a delightful guest, he was careful to observe the amenities in every way at all times. Greatness in an individual seemed to him to be made up of the little things. His wide interests and genuinely tolerant attitude made it easy to discuss with him things which in most circles would be matters of controversy.

He was married to Miss Emma Lay of New York on April 10, 1910. His devotion to his home was exemplary. He is survived by his wife and four boys.

A. E. T.

The trustees of the Teachers Insurance and Annuity Association announce the retirement on September 1 of Dr. Henry S. Pritchett, President of the Carnegie Foundation since 1905, and President of the Association since he coöperated with Mr. Carnegie in organizing it in 1918. Established by the Carnegie Corporation to supplement the work of the Carnegie Foundation by the provision at cost of annuity and insurance contracts for university and college teachers, the Association now has resources of twenty million dollars and has issued fifteen thousand contracts to officers and teachers in eight hundred universities, colleges, and research institutions.

The trustees have chosen as Dr. Pritchett's successor in the Presidency of the Association, Professor James W. Glover, Chairman of the Department of Mathematics of the University of Michigan. He was chairman of a committee of the American Institute of Actuaries which coöperated in inaugurating the Association and has been a trustee of the Association since its establishment. A graduate of the University of Michigan and of Harvard University and a Doctor of Philosophy of Harvard, Professor Glover organized at the University of Michigan in 1902 the courses in financial, statistical, and actuarial mathematics which have since developed, under his supervision, into the largest department of its kind in any university, having trained more than three hundred students who now hold responsible actuarial, statistical, and executive positions throughout the country. Professor Glover has served as Vice-President of the American Institute of Actuaries and of the American Statistical Association, as associate editor of this JOURNAL and as member of the councils of the American Mathematical Society and the Casualty Actuarial Society. He is the author of many well-known scientific publications concerning actuarial, insurance, and vital statistics; has been director and consultant of a large number of insurance and industrial companies; and has served the governments of Michigan, Wisconsin, the United States,

[1] The University of Minnesota is now preparing a memorial volume to Dr. J. Arthur Harris. It will contain several of his recent unpublished works and a complete bibliography of his writings.

and Canada in census, labor, insurance, pension, and banking legislation, notably in preparing the United States Life Tables for the Bureau of the Census and as a member of the Advisory Board in organizing the United States Bureau of War Risk Insurance.

Mr. Edgar Sydenstricker, Director Research Division, Milbank Memorial Fund, has returned from a visit to Geneva where he sat with the Health Committee of the League of Nations in connection with its deliberations on the proposed collaboration of the International Health Organization of the League of Nations with the Ministry of Health in China. From Geneva he went to China where he was asked to serve as a technical adviser on the rural health program at Ting Hsien which the Chinese Association of the Mass Education Movement is conducting with the aid of the Milbank Foundation. He also visited Nanking for the purpose of giving a series of lectures at Nanking University on vital statistics and population problems for the economic and agricultural faculties and members of the Central Statistical Bureau and the Legislative Yuan of the National Government and advised with the Ministry of Health on epidemiological and statistical matters.

Dr. W. Randolph Burgess, formerly Assistant Federal Reserve Agent at the Federal Reserve Bank of New York, was appointed Deputy Governor of the bank, effective May 1.

Dr. Ralph J. Watkins has resigned from the Research Staff of the National Bureau of Economic Research to accept the position of Director of the Bureau of Business Research and Professor of Statistics in the University of Pittsburgh. He will take up his duties in Pittsburgh on July 1.

Dr. W. L. Crum, who had been Editor of the *Review of Economic Statistics*, became Director of the Economic Service of the Harvard Economic Society on June 1, 1930.

Dr. Edwin Frickey became Editor of the *Review of Economic Statistics* on June 1, 1930.

Dr. George A. Lundberg has been appointed Director of the Research Bureau of the Federation of Social Agencies of Pittsburgh, which was recently established under a three-year grant from the Buhl Foundation. He has been a member of the Department of Sociology at the University of Minnesota during the past year and was formerly at the University of Pittsburgh.

Recent accessions to the statistical staff of the Committee on Cost of Medical Care are Dr. Isidor S. Falk and Dr. Maurice Leven. Dr. Falk is Associate Director of the Study.

Dr. Warren C. Waite, who has been with the Bureau of Agricultural Economics since September 1, 1928, resigned July 15 to return to the University of Minnesota to continue teaching and research work in Agricultural Economics. While with the Bureau Dr. Waite has made studies of consumer demand for dairy products, by questionnaires, for rural and urban areas in New Jersey and by means of sales in selected chain stores in New York City. The results of these studies will be published later by the New Jersey Experiment Station and the Bureau of Agricultural Economics.

Dr. Edmund E. Vial has been appointed as Senior Agricultural Economist in the Bureau of Agricultural Economics, Division of Statistical and Historical Research. He will conduct research studies in agricultural price analysis. Dr. Vial is a graduate of the University of Illinois and received his M.S. and Ph.D. at Cornell University, where he taught in the Summer School, 1930. Since September, 1927, he has been employed as statistician by the N. V. Potash Export Company of New York City. Dr. Vial expects to take up his duties in the Bureau at the close of the Summer School.

Miss Anna Michener, recently of the Statistical Staff of the National Bank of Commerce in New York, is now Economic Statistician for the Federal Farm Board in Washington.

ADDITIONAL COMMITTEE APPOINTMENTS FOR 1930

Committee on Social Statistics

Meredith B. Givens
Leon C. Marshall

MEMBERS ADDED SINCE JUNE, 1930

Anderson, Edward H., Professor of Business Administration, Ouachita College, Arkadelphia, Ark.

Anderson, Harold J., Student, School of Business, Columbia University, New York City

Ayres, Dr. Edith, Department of Economics, Washington Square College, Washington Square East, New York City

Bellows, Marjorie I., Division of Vital Statistics, New York State Health Department, Albany, N. Y.

Breithaupt, L. R., Oregon State College, Corvallis, Ore.

Brownlee, R. Bruce, Manager, Research Department, Standard Statistics Company, Inc., 200 Varick Street, New York City

Burch, Guy I., Population Reference Bureau, Washington Square East, New York City

Carr, Charlotte E., Industrial Consultant, Charity Organization Society of New York, 105 East 22 Street, New York City

Clark, Stanley, Statistician, B. C. Electric Railway Company, Ltd., Vancouver, Canada

Cohenour, Vincent, Research Assistant, University of Chicago, C. & A. Building, Chicago, Ill.

Croake, Richard P., New York Telephone Company, 140 West Street, Room 1725, New York City

Dennis, Roy E., Treasurer, Dennis Lumber Company, 823 Michigan Trust Building, Grand Rapids, Mich.

Fick, Knud, American Telephone and Telegraph Company, 195 Broadway, Room 750, New York City

Flatt, Leon B., University Press, Austin, Texas

Foster, Freling, 333 East 68 Street, New York City

Franseen, Arnold L., Student, Tripp Hall, University of Wisconsin, Madison, Wis.

Gill, John D., The Atlantic Refining Company, 260 South Broad Street, Philadelphia, Pa.

Grant, Albert, Psychological Laboratory, Board of Education, Cincinnati Public Schools, 216 East 9 Street, Cincinnati, Ohio

Halpert, Walter, Stein, Hall & Company, Inc., 285 Madison Avenue, New York City

Ham, Helene E., Economic Research, Federal Reserve Bank of New York, 33 Liberty Street, New York City

Henry, Malcolm H., Mathematics Department, Michigan State College, East Lansing, Mich.

Hoover, Edgar M., Jr., 22 Shepard Street, Cambridge, Mass.

Hopf, Harry A., H. A. Hopf and Company, Management Engineers, 40 Rector Street, New York City

Huang, Dr. Hsien-Ju, Assistant Professor of Economics, Yenching University, Peking, China

Huntington, Dr. Emily H., Department of Economics, University of California, Berkeley, Calif.

Johnston, Dr. J. A., Professor of Business Administration, University of Oregon, Eugene, Ore.

Jordan, Dr. Emil L., Statistician and Economist, International Public Service Corporation, 27 William Street, New York City

Long, Aubrey E., Bookbinders Trade Association, 15 Park Row, Room 921, New York City

Lurie, Sidney B., Theodore Prince & Company, 120 Broadway, New York City

MacNitt, Ruth, 809 East Washington Street, Ann Arbor, Mich.

Maverick, Lewis A., Student, University of California, Berkeley, Calif.

McGuire, Dr. Constantine E., Consultant in Economic Research, Cosmos Club, Washington, D. C.

Merkle, John W., United States Rubber Company, Detroit, Mich.

Merrill, Ray L., Statistician, Kraft-Phenix Cheese Corporation, 400 Rush Street, Chicago, Ill.

Neger, Elliott, Student, School of Commerce, New York University, 236 Wooster Street, New York City

Nelson, John, The National City Company, 52 Wall Street, New York City

O'Toole, Alphonsus L., Instructor in Mathematics, College of Engineering and Architecture, University of Michigan, Ann Arbor, Mich.

Ries, Henry F., Guaranty Life Insurance Company, Davenport, Iowa

Reimers, Alva R., 315 Cambridge Street, Grand Forks, N. D.

Rodda, Ruth, Arma, Kan.

Stephan, Frederick F., Assistant Professor of Sociology, University of Pittsburgh, Pittsburgh, Pa.

Storer, Robert W., Statistician, Tri-Continental Corporation, 54 Wall Street, New York City

Stugard, Erwin, Bond & Goodwin, Inc., 63 Wall Street, New York City

Swerdlow, Irving, 178 Orient Street, Bayonne, N. J.

Viner, Edward A., Head of Statistical Division, J. &. W. Seligman & Company, 54 Wall Street, New York City

Vogel, Harold A., School of Business, University of Minnesota, Minneapolis, Minn.

Vreeland, Wendell, Educational Research, Board of Education, 1354 Broadway, Detroit, Mich.

Walters, E. L., The Hills Brothers Company, 110 Washington Street, New York City

Ward, Charles F., Manager, Statistical Department, Appenzellar, Allen & Hill, 55 Broadway, New York City

Weintraub, David, Economic Research, Amalgamated Clothing Workers of America, 15 Union Square, New York City

Welsh, Lee M., Economic Research Department, Procter & Gamble Company, Gwynne Building, Cincinnati, Ohio

Wermel, Michael T., Manager of Accounting and Statistical Departments, Farmers' Agriculture Coöperative Marketing Association, 304 Greenwich Street, New York City

Werner, O. Helmut, Dubied Machinery Company, 139 Franklin Street, New York City

Weston, Janet L., 915 West Springfield, Champaign, Ill.

Willbach, Harry, New York State Committee on Mental Hygiene of the State Charities Aid Association, 215 Fourth Avenue, New York City

Younger, Gustavus A., Hartshorne, Fales & Company, 71 Broadway, New York City

REVIEWS

The National Income and Its Purchasing Power, by Willford Isbell King. New York: National Bureau of Economic Research. 1930. 394 pp.

This is an interesting and valuable addition to the studies of national income issued by the National Bureau of Economic Research. After three introductory chapters dealing with definitions of income, general character of estimates, the industrial classification of the population, and the nature of the index numbers used as "deflators," there follows a group of chapters (IV–VII) covering the industrial origin of income, the share of property and enterprise, the share of wage workers and salaried employees, and the developments with respect to concentration of income. A third group of chapters (VIII–X) deals particularly with income derived from corporate organizations and the gains and losses of investors, taking into account the factors of new investments and changes in market values of securities as well as dividends and interest. Chapters XI–XVIII involve a consideration of special aspects of income in agriculture, mining, manufacturing, construction, transportation, banking, government and merchandising. The material in the last chapter (XIX) relating to residential property and income from miscellaneous sources might well have been omitted or tucked in elsewhere.

The volume contains 133 tables and 60 charts, which indicates the large number of matters dealt with and the highly tabular and statistical character of the material. The book would be easier reading if many of the relatively unimportant tables were either eliminated or relegated to an appendix. From the reader's standpoint, further, the book would be much more effective had the author made an effort throughout to distinguish sharply between major and minor problems and conclusions. More care might also have been taken with definitions and assumptions and a more clear-cut statement of the purposes of the study would have been helpful.

As in other studies of the Bureau the emphasis is on "realized income" (income which does not include effects of changes in value of inventories or other property) from the standpoint of the individual recipient, but considerable attention is also directed to the problem of measuring "total book income" (which does include effects of value changes). As in the earlier studies, also, questionable methodology is employed in combining with total market income—the amount available to individuals during a given period to spend, wisely or foolishly, or to save—an estimate of the net use value of durable consumers' goods. In the last analysis, of course, our total realized income in a particular period is the value of the amount consumed in that period, including the services of durable goods, and the value of all services performed by individuals for themselves and families up to the very instant of consumption, but there seems to be little excuse for a calculation which takes a particular item from the total income so defined and adds it to an income total defined from an entirely different stand-

point. This procedure is especially objectionable in view of the fact that the value of the services of housewives—which can be estimated with much greater ease and accuracy than the net use value of residences and other durable consumers' goods—is excluded from the total income. The reviewer should admit, however, that the validity and usefulness of the figures presented are not appreciably impaired because of the procedure criticized, particularly since the total of the "imputed income" is not large relatively and is shown separately throughout.

Most accountants object to the definition of "book income" implied here, and it is true that in accounting practice generally value changes during the period under consideration are ignored in the determination of income. Incidentally accountants will do well to ponder the statement "what may happen in the future is of no concern whatever to the accountant or statistician estimating the income of an enterprise for a given period." (p. 38)

With respect to the industrial classification of the population the author shows, as would be expected, that the number of persons engaged in agriculture, railroading, mining, and construction, respectively, has shown relatively little change (ignoring the war period) over the last two decades, whereas there has been a decided increase in the number engaged in manufacturing, merchandising, banking, light and power enterprises, telephone and telegraph enterprises, governmental activities, and miscellaneous industries. This hardly means as the author states that "with increasing prosperity the people of the country have demanded less of these old-fashioned products and more of the newer types of goods and services," but it indicates that there has been a less marked increase in output of some of the basic industries than in the output of some of the less essential lines (although even this conclusion must be tempered with a recognition of the fact of increasing output per person engaged in many basic lines). The study shows, as would also be expected, that there has been a decidedly more rapid increase in the number of so-called salaried employees in industry in the past twenty years than in the number of wage workers. It also shows that the total number of individual entrepreneurs has not increased and that there has been a decided decrease in such entrepreneurs in manufacturing. Lawyers and other professional men, incidentally, are treated as entrepreneurs in the tabulation given.

In the statement of the reasons for the use of indices of prices of consumers' goods as deflators the author's assertion that "the present price of wheat is dependent upon the anticipated price of the bread into which it is ultimately to be converted" can hardly be stomached as good price theory by the economist. In connection with the tables of these indices on pages 68–69, one wonders why the January index is almost invariably higher than the July—except in the war years. Is this due to high fuel bills and doctor bills in the winter?

As indicated by the title, the author is particularly interested in measuring income trends and developments on the basis of purchasing power, and undoubtedly this is the most important aspect of the work. In this connection it is shown that the total income of 1928, in terms of purchasing power, is about 75 per cent greater than in 1909; that the percentage of such purchasing power going to property owners and entrepreneurs has decreased appreciably since 1909,

although remaining almost constant since the War; that "the average American was approximately one-third better off in 1927 than he was in 1921"; that the purchasing power of income drawn from manufacturing was about 75 per cent greater in 1923 than in 1909, although the share going to entrepreneurs and property owners in this field increased in the same period by only one-third; that "to the owners of American industries the Great War brought, in the main, severe losses instead of gains;" that the trend of the average annual earnings of the wage worker, in terms of purchasing power, has been upward for most of the period studied in virtually all lines, although the fluctuations have not been at all uniform; that the average real income of every class of salaried employee declined markedly during the War and was in 1927 in some lines slightly lower than in 1909; that the purchasing power of all common stock dividends had increased by 1928 to nearly two and one-half times the figure for 1920 and to a point nearly 20 per cent higher than the total in 1927, while the corresponding figures for preferred dividends and corporate interest disbursements, respectively, show much less marked fluctuations.

In Chapter IX estimates are made of the market values year by year of the property rights of individual security owners in industries dominated by corporations. These estimates show, as would be expected, highly fluctuating amounts from year to year, but they do not show, as might be expected, any noticeable upward trend in 1913 dollars, even in the manufacturing field, and in the case of railways the trend has been distinctly downward. It would appear from these figures that the success of some enterprises has been roughly matched by the failure of others, at least for the period 1912–1923. In this chapter are also estimated in terms of 1913 as well as current dollars the annual gains and losses to individuals of their property rights in industries dominated by corporations, with allowance being made for new money invested. In Chapter X (pp. 225–290) are presented the results of an elaborate inquiry into gains and losses of corporate investors (individuals) in certain selected fields based on changes in market values of securities held. It is to be doubted, however, if speculators can glean either comfort or a basis for recoupment from the figures presented. The principal value of this part of the study, in the reviewer's opinion, lies in the implicit showing of the underlying unreasonableness of stock market fluctuations and the questionable character of income (or loss) represented by fluctuations in security values from the standpoint of either the individual or the nation, and the advantage of the conception of realized income over that of total book income. Offhand it would seem that the work of converting all of the wildly fluctuating figures here displayed into 1913 dollars was not wisely spent. Of special interest to accountants in this chapter is the chart on page 284 which makes a vivid comparison of the trend of the "logical" value of stocks and that of the "actual" value for railway, switching, and terminal companies. This bears out—for the particular field—the proposition often urged in stock market circles in recent years to the effect that the fixed property accounts shown on the balance sheet bear no relation whatever to the commercial value of the enterprise. It looks very much as if in some industries corporate saving was a case of throwing good money after bad from the investor's standpoint.

In the last group of chapters are brought out many important (and some unimportant) conditions and tendencies in agriculture and other fields. In these sections quantitative expression is given to the relatively unsatisfactory position of the farmer in recent years, to the tendency toward domination of the mining field by a few large companies, to the rapid development of the light and power and telephone industries, to the growth of governmental activities, and to trends in certain branches of merchandising.

The style of the book—perhaps unavoidably in view of its nature—is rather tiresome, and it is not entirely free from careless statements. For example, the reference to Table XXIV on page 182 should include mention of Table XXXV, and the statement at the bottom of page 185 relative to the position of bondholders should explicitly take into account the fact that the total investment of the bondholder has increased for the period under consideration.

W. A. Paton

University of Michigan

Women in Gainful Occupations, 1870 to 1920. A Study of the Trend of Recent Changes in the Numbers, Occupational Distribution and Family Relationship of Women reported in the Census as Following a Gainful Occupation, by Joseph A. Hill. Washington: United States Census Monograph IX, 1929. 416 pp.

This treatise, the second since 1900 on this subject from the Census Bureau, sets forth the data analyzed three ways: first, the actual numbers of women reported as gainfully employed in each census year since 1870 with their various characteristics; second, comparisons of the changes from one census to the next, particularly those between 1900 and 1920, in the proportions within the group of gainfully employed women, of those with the various characteristics and in the different types of work; and thirdly, the proportions in the important demographic groupings of women who were reported as gainfully employed. "Women" means females sixteen years of age and over.

For those carrying on research on the subject of employment and women's relation to it, this book will be a rich source of indispensable material. Its style is that usually found in Census Bureau publications. The interpretative text is benevolently masculine.

The number and per cent of women following gainful non-agricultural pursuits rose from 1,321,364, or 11.8 per cent of all women, in 1870, to 7,306,844, or 21.3 per cent, in 1920 (Table 12). When this later rate was "adjusted" to make allowance for the shift from 1910 to 1920 in the relative strength of the various age groups, it became 22.1 per cent (Table 17).

Owing to several disturbing factors, the data on women in agricultural pursuits are not satisfactory. As reported, they constituted 2.9 per cent of the women in 1870, 3.1 per cent in 1880, 3.1 per cent in 1890, 3.3 per cent in 1900, 4.8 per cent in 1910, and 2.6 per cent in 1920. If one agrees that it is important to distinguish sharply between indoor and outdoor work of women on farms, one can sympathize with the Census Bureau in its efforts to draw that line consist-

ently in each census. The author explains that in the instructions for the 1910 census, emphasis was put on returning as a farm laborer every woman who worked regularly at outdoor farm work. In 1920 the emphasis was put on not returning as a farm laborer a woman who worked outdoors only a short time each day or only occasionally. She had to work "regularly and most of the time" at such work to be considered as gainfully employed. "This change of emphasis came about because a study of the occupation returns convinced those who had charge of the tabulation in 1910 that many women had been returned as farm laborers who could not properly be regarded as such—that there was, in short, an overenumeration of women farm laborers." There were also other differences in definition which tended to narrow it still further and thus to render incomparable the figures in 1900, 1910, and 1920.

Into the census category of women following agricultural pursuits were put 896,057 women. A total of 3,581,158 agricultural workers were reported for pursuits in which women furnished at least 5 per cent of the total workers (Table 43). Of the women agricultural workers, 403,009 were farm laborers on the home farm, 198,979 were farm laborers working out, 9,984 were other agricultural laborers, 256,160 were farmers, dairy farmers and stock raisers, and 27,925 were reported as following other agricultural pursuits. Reluctant as one may be to add to the perplexities of the occupation classification or to raise a feminist hue and cry, one cannot believe that women's contribution to agriculture or to the operation of farming is truly reflected by these figures. Care of animals—cows, pigs, poultry, bees—gardening and raising of fruits, preservation of meats and other foods, peculiar to the farm economy, boarding of men farm hands (of which the census reported 1,792,507!) and incidental help with the outdoors work all go into a job essential to the operation of the farm itself and differing widely from the occupation of the housewife in the city. The picture of the workers in the farm industry will not be complete without some accounting for the services of what might be termed "able-bodied wife of farmer." Like the able-bodied seaman, she has no exact layout of duties, but as he is putting necessary energy into the basic industry of shipping, so she is putting necessary energy into the basic industry of farming, and not simply keeping house for her husband and attending to her children.

Complaints have been made elsewhere of the Census Bureau's definition of gainful employment as applied to women. By requiring that the woman must receive a definite money return or its equivalent, it throws into the category of housewives or of persons with no occupation, those who spend the major portion of their time in unpaid service, in the enterprises of their husbands, fathers, sons, and other relatives. Their services are real and tangible and lie in the field of economic competition, since they would need to be replaced if withdrawn. Work of this sort differs from housekeeping in that it can be evaluated, and in that it often takes women out of their homes just as regularly as if they were paid wages. Were it possible to take account of these services, women's contribution to trade and industry might be found to be somewhat different from that shown by census figures.

The limits of this review do not permit a critical examination of more than two

or three of the many points brought out in this extremely useful package of facts. While the 1920 data are examined for the extent of gainful employment of women of different ages, the discussion as to trends centers around the varying proportions of all employed women who were in the age groups of 16 to 24, 25 to 44, 45 to 64, and 65 years and over, as shown in the different census years (Table 14). From 1910 to 1920 a decreasing percentage of occupied women were in the youngest age group and an increasing proportion in the other groups, including both of the older ones. But tucked away in the middle of a voluminous set of detailed tables in the appendix are figures (Tables 144 to 148) which tell a tale about the proportions of each age group of each nativity who were gainfully employed in 1900 and 1920. The per cent employed of all women aged 16 to 24 increased from 31.6 to 37.6; of all women aged 25 to 44, from 18.1 to 22.4; of all women 45 to 64, from 14.1 to 17.1; but in the group 65 years and over it declined from 9.1 to 8.0. In the New England and Middle Atlantic states, with the exception of New Jersey, and in six other scattering states, the proportion in this last age group increased slightly from 1900 to 1920. The decline in employment of these older women occurs in all of the nativity groups except that of native white of foreign or mixed parentage. It is possible that it is related to the facts, elsewhere recorded, that there was an actual decline in the number of women servants, dressmakers and seamstresses (not in a factory), laundresses (not in a laundry), milliners and millinery dealers, and boarding and lodging house keepers. The testimony of the 1930 census figures on the extent of employment of these older women is awaited with interest.

Comparisons of the per cent of women gainfully employed in 1920 in non-agricultural pursuits in the different states shows highs in the District of Columbia and Rhode Island of 49.6 and 35.6 per cent, respectively, and lows in Arkansas and Mississippi of 9.6 and 11.4 per cent (Table 20). In general, gainful employment, including agriculture, appears to the extent to 24.3 per cent of the women in the north, 23.9 per cent in the south, and 21.3 per cent in the west (Table 3). It is probable that a more adequate accounting for women's agricultural services would increase the percentages and might change the relative positions of the sections.

The gainful employment of married women receives detailed discussion. Those in non-agricultural pursuits constituted 3.3 per cent of the married women in 1890, 3.9 per cent in 1900, 6.8 per cent in 1910, and 7.3 per cent in 1920 (Table 48). The percentage of increase in the employment of married women exceeds the increase in the employment of unmarried women in each of the three decades between 1890 and 1920 (Table 49). Among the 11,187,355 native white married women of native parentage, 6.3 per cent were reported in 1920 as gainfully employed. Similar proportions were found among other white groups. Among the 2,036,419 Negro married women, 32.5 per cent were gainfully employed (Table 51). Among all white groups the proportion of gainfully employed decreases consistently in the age groups over the 16 to 19 year old group. With the Negroes this proportion consistently increases and reaches its maximum in the age period 35 to 44 years (Table 53).

The explanation of this peculiar divergence lies outside the knowledge provided by the census, but it is a situation well worth study.

The trend in type of employment of married, as of other, women was away from being milliners, dressmakers, laundresses (not in a laundry) and keeping boarders and lodgers. The large increases were in skilled, semi-skilled, and unskilled factory work, in clerical service, including stenographic service and bookkeeping, in employment as saleswomen, waitresses, hairdressers and manicurists, and as nurses and teachers (Table 57).

Painstaking attention is devoted to data for 1910 and 1920 showing trends in the employment of women of the several nativity groups.

The data on family relationships of gainfully employed women relate to 1900 and 1920, and come from special tabulations for 373,204 gainfully employed women in eleven selected cities—11.7 per cent of all gainfully employed women in cities with a 1920 population of 100,000 and over.

In 1900, 69.2 per cent of gainfully employed women in these cities were single, 13.4 married, and 17.4 widowed and divorced. In 1920, the proportions were 60.1, single; 22.9, married; and 16.9, widowed or divorced (Table 98).

Since 1900 the proportion of those living at home increased from 71.2 to 78.6 per cent (Table 97). In those two decades the proportion of single women with gainful employment increased 52.5 per cent, the ratio of married women with jobs increased 199.7 per cent, and of widows and divorced women, 70.6 per cent (Table 98). In each one of the eleven cities the gain in the percentages of employed married women exceeded the gain in employment among the other two groups (Table 99).

In the case of 27.2 per cent of 293,213 gainfully employed women, there were no dependent members of the family, that is, either the woman was the only member of the family (7.6 per cent), or all of the other members were gainfully employed (Table 109). In 7.5 per cent of the cases the woman was the only gainfully employed member of the family and there were one or more dependents. In 3,519 cases (1.2 per cent) the woman was the only gainfully employed person in a family of four or more persons. Twelve other categories appear, in which there are two or more employed persons and one or more dependents. In 11.2 per cent of the cases there were two employed persons and a third member of the family group unemployed (Table 109).

If and when the Census Bureau tabulates the census data on families in other communities, it will be possible to secure a much larger body of facts on the family relationships of employed women.

The arrangement and treatment of a great mass of items of information, such as were marshalled in this volume, is at best a difficult and laborious task. There is no one best way to do the whole job. But if one accepts the criteria of clearness, of systematic arrangement and of accuracy, he will classify this as an excellent piece of work.

NEVA R. DEARDORFF

New York City

Earnings of Factory Workers, 1899 to 1927, by Paul F. Brissenden. Washington: United States Census Monograph X. 1929. 424 pp.

The task of examining the reports of the Census of Manufactures with the aim of finding out what light they might throw upon the earnings of factory workers over a period of years was Dr. Brissenden's. As a result of his efforts and that of the staff of the Census Bureau, we have a most welcome addition to that very serviceable group of studies—the census monographs.

The two statistical items at the basis of the investigation are average number of wage earners and total payrolls. When payrolls are divided by the average number of workers the "census average wage" is computed. This ratio has been generally employed to record changes in money and real earnings. Although the limitations of the measure have been recognized for a long time, it has never received such a rounded discussion as is given by the author in what is perhaps one of the best chapters of the book. The census average wage registers neither full-time earnings nor actual earnings of those who are usually considered to be attached to industry. Comparisons among industries are influenced by changes in the relative proportion of male and female workers and of skilled and unskilled labor. Then, too, the average number of workers reported on the payroll is somewhat larger than the facts warrant.

How does Dr. Brissenden go about improving upon the census average wage? It must be remembered that the primary concern of the monograph is to show changes in per-capita money and real earnings as well as to present the best approximations of the actual sums received. To achieve these aims, the aid of materials other than those derived from the census is utilized. And the best attack is found to be a circuitous one as is shown by the following outline of the procedure.

The average weekly earnings of factory workers in the "busiest week" of the year 1904 (which was, perhaps, not so busy) were taken as starting points. Full-time wages for that same week were estimated on the basis of the ratio of actual employment to full employment of trade unionists in New York State and some additional adjustments supplied by figures from the United States Census of Occupations. For the other census years, full-time weekly wages were obtained by using changes in the census average wage to represent changes in full-time earnings. The weekly figures were turned into annual earnings by multiplying by 51 weeks.

The derivation of full-time earnings is only an intermediary device for getting the actual money and real wages of workers "normally" attached to manufacturing industries. The numbers out of work had to be taken into account. To correct for unemployment, first, indices of employment were constructed by the novel method of deriving them from indices of production. Then, ratios of actual to full employment were computed on the basis of an average of two estimates both of which were taken essentially from W. I. King's *Employment Hours and Earnings in Prosperity and Depression*. Full-time yearly earnings were multiplied by the ratios of actual to full employment to get actual earnings per worker. Indices of actual earnings were then divided by cost of living indices to show changes in purchasing power of these incomes. Absolute amounts

were estimated and indices constructed for intercensus years by interpolating on the basis of information from state bureaus.

A section on variability of earnings in 1919 is presented and comparisons are made with estimates of variability for the years 1899 and 1904. There is also a chapter on average hours worked per week and average hourly earnings.

Students of wage statistics will undoubtedly take issue with the author on several questions pertaining to methods employed. The reviewer has found particularly troublesome the concept of full-time earnings—a measure that has had a greater influence upon the final results than appears to have been realized by the author.

To convert average weekly earnings to a full-time basis by the methods outlined above is to proceed upon the assumption that the percentage of unemployment approximates the percentage that actual time worked is to full time. For example, on page 297, $10.06 was the average weekly earnings of factory workers in 1904. This sum is divided by .87 (the ratio of actual to full employment) and $11.56 is obtained as the estimated full-time wages. Now this difference of $1.50 is the additional amount which those on the payroll would have received had they worked full time; or it might be said to be the extra sum which all attached to industry—those on and those off the payroll—would have received if they all had worked full time. Stated in another way, if only 87 per cent were employed, it also represents 87 per cent of a full-time week of those on the payroll.

The correctness of this assumption is very doubtful. King for instance, in his study on *Employment Hours and Earnings*, showed that the decline of the percentage of full time worked by factory workers was 6.5 per cent from the peak to the trough in the first post-war business cycle. On the other hand, the decline in the number of employees on the payrolls was 25.59 per cent from peak to trough. This evidence would seem to point to the conclusion that the percentage of unemployed is not a representative measure of the percentage of time lost as compared with nominal full time.

Another difficulty is to understand why the census average wage is considered to be a good measure of the changes in full-time weekly or yearly earnings. There may be a clue on page 282 when it is stated that the margin by which the census average wage falls short of full-time earnings is constant. This, however, is begging the question since the constancy between the two averages is the direct result of the method of computing full-time weekly earnings, as can be seen on page 297 where $11.56 was multiplied by relatives derived from the census average wage. The conclusion seems to be unmistakable—changes in *full-time earnings* are directly dependent upon changes in the census average, and to a lesser extent changes in *actual* earnings are also influenced by the census average wage in view of the fact that actual earnings are derived from full-time earnings.

These two sceptical comments are submitted for inclusion in the fairly large list of doubts which Dr. Brissenden has himself been so very careful to point out throughout the pages of this thoroughgoing piece of work.

ASHER ACHINSTEIN

New York State Board of Housing

Die Statistische Analyse der Konjunkturschwingungen, Probleme der Weltwirtschaft, by Walter Hahn. No. 47, Jena. 1929. 200 pp.

This book discusses the methods of applied mathematics and statistics in reference to the analysis of cyclical fluctuations. The problems of elimination of secular movements and of seasonal variations are explicitly ruled out of the discussion, with the assumption, however, that such elimination is possible. One wonders then how such an extensive mathematical treatment of business cycle swings alone is possible, particularly since the more complex methods are relegated to an appendix. After reading this book, one might still question the relevance of a large part of the discussion.

The text is divided into two main sections: (1) the analysis of *absolute* movements, i.e., of cyclical changes in each time series taken by itself, and (2) the analysis of *relative* movements, i.e., of cyclical changes in one time series as compared with another. In the first section (pp. 16–63) the topics are: increment and growth—in the succession of cyclical phases; wave length and amplitude—in the successive cyclical swings. But of these two topics, the former, that of increment and growth, occupies the bulk of the first section. In the second section (pp. 64–136) the topics are: comparative intensity and timing of cyclical swings in two or more series; comparative changes in the direction of cyclical movements. The bulk of this section is devoted to correlation, with a final chapter dealing with moving correlation.

The emphasis of the discussion is almost exclusively on the mathematical aspects of analysis and comparison. The complexities encountered in concrete problems of cyclical study are disregarded to such an extent that in reading the book the question of relevance arises time and again. Thus the first section is primarily an exposition of the elements of differential calculus as a method of studying increments, and of logarithms as a method of studying growth. The importance of these mathematical elements in the analysis of cyclical swings or of cyclical phases is not quite clear, and little light is shed in the chapter which deals with the application of the mathematical discussion to time series. In this chapter one finds a treatment of the applicability of continuous-function-analysis to discontinuous time series, of the use of logs and first differences in measuring growth, and computation of the duration of a fairly repetitive cyclical swing. But little space is devoted to simple and possibly more effective methods of studying cyclical swings as they occur in reality, little appreciation is given of the difficulties encountered in concrete analysis, a failure not remedied by a set of illustrative charts given as a supplement at the end of the book.

This over-emphasis of the mathematical part of the discussion is not so apparent in the second section of the book, which deals primarily with correlation, a method whose mathematical foundation is of more direct relevance to the analysis of cyclical swings than is that of differential increments or of logarithms. Of interest are the chapters devoted to the scatter about the regression lines, and especially the discussion of the correlation surface for sinus curves, a surface which should be representative of the correlation of time series. In this chapter and also in the mathematical appendix (pp. 136–200), a competent reader might

find interesting suggestions and references to Continental literature not familiar to the English-speaking student.

In the field of business cycle study, little positive work was done on statistical method for analyzing successive cyclical swings in each series proper, little as compared with problems of eliminating trends or seasonals or of correlation of cyclical changes in two or more series. The establishment of the average cyclical swings and of their characteristics has not been attempted frequently. The usual Harvard type of analysis does not go beyond the plotting of the cyclical residuals and the computation of the standard deviation and of the lag (for correlation), while the Fourier type of analysis presents serious limitations peculiar to itself. The whole empirical and logical problem of average attributes of cyclical swings found in time series has hardly been touched, and such subsidiary questions as the elimination of random changes and the discounting of the influence of secular movements on cyclical averages have not been fully solved. The real task of a general treatise on the statistical analysis of cyclical swings, whether in this country or abroad, is the discussion of these questions in the light of the mathematical and logical nature of potential and available methods on one hand, and of the concrete problems presented by various large groups of time series on the other. Dr. Hahn's discussion, by and large, has overlooked this problem, and consequently contributes but little to the elucidation of the essential questions involved.

SIMON KUZNETS

National Bureau of Economic Research

Základy Teorie Statistické Metody, by Stanislaw Kohn. Prague. 1929.

The book under review is published in the Czech language and edited by the Central Statistical Office of Czechoslovakia. The author is a Russian, however, a pupil of the well-known Russian statistician, Professor A. A. Tschuprow, with whom he was closely associated until the untimely death of the latter in 1926.

Professor Tschuprow died without having the opportunity to present in one systematic book his very important recent findings in the mathematical theory of statistics. During the last ten years of his life he published a great many articles in various languages in European statistical periodicals, but only his study[1] on the theory of correlation was ever published in book form (in German and in Russian). Mr. Kohn, in *The Theory of Statistical Method*, gathers many of these findings together and presents them as a part of his systematic development of the statistical method. For this one reason, if for no other, the book would be of special interest to American readers.

It should not be thought, however, that Mr. Kohn's book is merely the presentation of Professor Tschuprow's findings. The author makes use of the enormous literature on mathematical statistics in English, French, German, Russian and Italian. Furthermore, on some important problems, Mr. Kohn differs from his teacher; for instance, in the philosophical interpretation of probability. Although both Mr. Kohn and Mr. Tschuprow subscribe to an objective interpreta-

[1] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelations Theorie*, Teubner, Leipzig, 1925.

tion of probability, as opposed to the subjective interpretation of the French mathematician, La Place, Mr. Kohn follows rather the German philosopher, J. Kries, and the English mathematician, R. L. Ellis, while Tschuprow follows more the French mathematician, A. Cournot.

Mr. Kohn continues the work of Professor Tschuprow in bringing together the mathematical and logical analysis of the a priori probability of the French, Russian, and German mathematicians and the empirical analysis of the English statistical school (K. Pearson and others). It may be considered as one of the principal achievements of Professor Tschuprow that he built a base of a priori conceptions for the empirical, a posteriori system of the English school of statistics, traditional also for America. And on this point Mr. Kohn presents, in clear statements and in comparatively elementary form, the principal theories developed by continental European mathematicians and statisticians with which American and English statisticians are not always acquainted. The table of contents of the book shows in itself how different it is from the traditional American book on the principles of statistics.

The second part of this book opens with an analysis of the difference between functional and loose (stochastic) causal relationship. The author connects the loose relationship with the conception of probability, the mathematical and logical analysis of which he has very largely developed in this part of his book. He gives considerable space to the exposition of the law of great numbers mainly as it was developed by the mathematicians of the French and Russian schools. The law of great numbers, in the case of quantitative variables (Chapter X, pp. 198–226), Mr. Kohn explains by applying the method of mathematical expectation, so extensively used by the Russian school of mathematicians and the principal mathematical tool used by Professor Tschuprow in his statistical analysis. Mr. Kohn, however, does not limit his presentation of the law of great numbers to mathematical analysis; he also gives the logical and philosophical conception of the law and its relationship to actual life (see especially Chapter XI).

The presentation of the law of great numbers is accompanied by a special chapter on the stability of statistical numbers, a problem very extensively discussed by continental European statisticians. The measure of the stability of statistical series, introduced by Professor Lexis and developed by Professor Bortkevicz and Professor Tschuprow, is discussed in Chapter XI, and this chapter, therefore, will be of especial value to those interested in the analysis of time series.

In the presentation of the theory of correlation, the author gives first the a priori notions and formulae and later connects them with the usual empirical formulae for correlation measurement. Chapter XIII is devoted to the analysis of the dependence between attributes. In this he may be said to follow Yule; his analysis, however, is different from that of Yule. He starts with the fundamental a priori criterion of dependence and passes later to the empirical characteristics.

In his development of the theory of correlation between variables, or better, of the theory of stochastic (collective) dependence between variables (Chapter XIV), Mr. Kohn follows the a priori theory developed by Tschuprow. For

pedagogical reasons, however, he develops the correlation theory in a different way, especially in order to show more clearly the relationship between the coefficient of correlation and the correlation ratio. This chapter may be considered as one of the most logically developed chapters in the book, yet intelligible to readers with the knowledge of only elementary mathematics.

Mr. Kohn pays attention, also, to the analysis of time series (Chapter XVI). He treats this problem in controversial manner—after presenting the traditional American method, he sets forth also the criticisms of this method, in particular by the Russian statistician, O. N. Anderson, one of the initiators of the variate difference method.

The theory of index numbers also is presented in controversial form. This chapter (Chapter VI) should be of great interest to American readers because considerable attention is given to American authors (I. Fisher, W. C. Mitchell, W. I. King), and to the theories developed by the English and German statisticians, especially by Professor Bortkevitz ("Zweck and Struktur einer Preisindexzahl," *Nordisk Statistisk Tidskrift*, 1923, 1924). Mr. Kohn criticizes particularly the disapproval of American statisticians toward the unweighted index number. He shows that this disapproval sometimes is based on mathematical misunderstanding (for instance, by F. C. Mills). (See footnote on page 108 of the book of Mr. Kohn.)

The thirty-five pages allotted to the bibliographical appendix give a very good review of the literature of the subject. Especially of interest is the fact that the literary index includes Italian, Russian, French and German publications which are not well known to American and English readers.

V. P. Timoshenko

Corporate Earning Power, by William Leonard Crum. California: Stanford University Press. 1929. 342 pp.

This book presents the results of what must have been an extremely laborious study of the federal statistics of corporate income, from 1916 to 1927, inclusive. Since we do not know so much about profits as about many other less critical economic phenomena, it is a useful pioneer work.

The nature of the study was apparently determined chiefly by the characteristics and limitations of the available material. It begins with an elaboration of what the author terms "profit ratios," i.e., the ratio between net income (less income taxes) and gross income. This ratio is computed, in all the years, for all industries combined and for major industrial divisions such as manufacture, trade, mining, service, construction and agriculture. The ratio is given for the main manufacturing groups from 1916 to 1926 or 1927, inclusive, and for selected groups and sub-groups in the years 1916 and 1926. There follows a section on the "earnings ratio" in selected years for these same classifications, the author defining as the earnings ratio the percentage of net income to total assets as carried in the reported balance sheets. Minor measures such as the rate of profits to gross sales follow. The author makes use of the balance sheet material more recently available to analyze the ratios of current assets to current liabilities, of net worth to total assets, of long term debt to fixed assets. In conclu-

sion, he turns his attention to the distinction of profitable from unprofitable corporations, according to industrial or size group.

What can be learned from material of this sort? Much less than would seem possible without a knowledge of the statistical limitations. Take, for instance, the "profit ratio." "Gross income," in the reports of the Bureau of Internal Revenue, includes all receipts except tax-exempt interest (ordinarily a very small factor). This is far from a desirable figure to use, for precise purposes. It embraces income from outside investments as well as from operations in the business in question, and also profits from "other operations," which are really net rather than gross. Net income, on the other hand, excludes dividends received—which are a net item, and should be in this category if the ratio is not to be limited to the operations proper of the company or industry. But these two figures are employed because better ones are not available throughout.

As the author points out, the "profit ratio" in itself is nearly useless for the comparison of the profitableness of different industries, since one industry may do a larger volume of business on a given amount of invested capital than another. It likewise throws little light on geographical comparisons of profitableness, since industries with one characteristic type of profit ratio are more prevalent in one region than another. The figures are not particularly significant for banking and public utilities, because of inestimable deficiencies in the figures for gross income in these cases. In farming they are not of much relevance because such a small part of agriculture is conducted by corporations.

The book does place some stress, however, on their meaning in tracing the fortunes of a single industry or group of industries over the period in question. Even here, the figure must be broken down into its component gross and net before we can know much about what has happened. A stable profit ratio with expanding business might mean something quite different from the same ratio with contracting business. The figures do show that 1916 was an extraordinarily profitable year, that in 1921 average profit disappeared, that recovery took place in 1922, and that from 1923 to 1926 profit ratios were stabilized somewhat below the 1916 level. These are scarcely startling discoveries. The figures also indicate that the difference between profit ratios in 1916 and in 1923–26 was due primarily to an expansion of gross rather than to a contraction of net. They show that gross fluctuates from year to year much less than net; that frequently a contraction in gross, or even a failure of gross to increase substantially, is accompanied by a larger decrease in net, and that sometimes net falls for a brief period while gross is enlarging. Reasonable inferences may be drawn from the figures, confirming knowledge we already have, such as the depression of coal and textiles, the rapid expansion of service industries and public utilities, the building boom. More detailed conclusions the reviewer would be inclined to question more closely than does the author; if wholly reliable, they would have to rest on further analysis of the movements of gross and net to determine the effects of relative price changes for products and materials, of changed methods and technique, of changed organization of industry. The author seems to assume too readily that 1916 can be taken as a typical pre-war year. And he fails to point out the rather obvious fact that, since 1923, the figures in most cases fur-

nish little support for the prevalent belief that industry has been reducing its profit margin in order to obtain larger sales.

Still stronger reservations must be made to the study of the ratios of return to total assets. Accounting methods are not uniform or sound enough to lead us to believe that the figures for "total assets" have anything like the same meaning in different industries, or indeed any meaning at all in many cases, beyond the capitalization of earning power. One is not surprised to learn that, in 1926, the ratio of return to assets was larger in industries characterized by closely held corporations like printing or stone, clay, and glass, than in industries whose securities are widely marketed like public utilities, since in the former case there have been fewer write-ups of original cash investment. And one is disinclined to see any great significance in the author's statement that, in the supposedly profitable year, 1926, the percentage of manufacturing earnings to "invested capital" was no larger than the average yield on long-term bonds of 4½ per cent. We should have to know more about the exact meaning of the figures called "invested capital" before admitting that Mr. Crum is correct in calling this a "generally unfavorable situation."

This difficulty of drawing really significant and reliable conclusions is largely due to the refractory nature of the statistical material itself, but it is also in part due to a prevailing method of handling statistical work exemplified by this book. Mr. Crum throws all the figures into the hopper to see what will come out. Break up enough crude ore, he seems to assume, and you will find whatever truth there is in it after you have done so. There is no preliminary setting up of a series of hypotheses to be tested, or of exact questions to be answered. There is not enough systematic preparation to discover what a given set of figures may possibly tell us. The result is that the reader is not presented with a well-organized method of assessing the material. If he exercises his critical faculty, he must continually question the logical usefulness of the presentation; if he does not exercise it, he is almost certain to draw hasty and unsound conclusions. Some of these, unfortunately, are shared by the author, in spite of the fact that no technical statistical errors may be found. The book would have been benefitted greatly, for example, by a preliminary inquiry into the possible economic significance of figures on the ratio of net to gross income, which would, in a thorough manner, have outlined the various implications theoretically involved in the possible variations of these factors.

Perhaps, however, this is to ask too much of a pioneer investigation, which after all involved a prodigious amount of labor and does break new ground.

GEORGE SOULE

New York City

Money Credit and Prices, by Jan St. Lewinski. London: P. S. King and Son. 1929. 144 pp.

This little book displays a combination of qualities that is all too rare in authors, a wide knowledge of pertinent literature, a keen theoretical mind, and a propensity for testing hypotheses by a range of facts that includes many lands

and several centuries. It offers little that is new in separate doctrines: the originality is rather in their synthesis. The author's thesis is that J. S. Mill and the contemporary banking school found truths that advocates of a managed currency nearly a century later would do well to heed. (The living Englishmen to whom this thesis applies are not named.) Fullarton, he tells us, "saw that no automatic devices can perform the delicate and arduous duties of the Directors of the [Central] Bank . . . in this respect as in many others his ideas are completely in harmony with those expressed by the Federal Reserve Board" in its Tenth Annual Report.

The first part of his book is a critique of the quantity theory, which he finds has many forms. Some of his points against the theory follow: P includes security and real estate prices. Taxes, gifts, and security clearing house settlements by offset do not fit into the equation of exchange. Increased (or decreased) M may sometimes so decrease (or increase) V as to cause a price decline (or rise). Changes in P may cause changes in V. Hoards, as noted by Fullarton, and bank reserves may cushion the effects of gold movements. The quantity of book credit (M'), as Mill recognized, depends not on the quantity of coin and bank notes, but upon *expected prices*. Increased gold production may increase T, as Hume pointed out. "It is easier to adopt [sic] the quantity of money and credit to a changed price level, than to modify prices by managing the circulation." Contraction of needed bank circulation will force a resort to less safe and convenient substitutes, as Fullarton states. It is easiest to effect credit control at a time of undue expansion. In a quiescent state a high bank rate may raise costs and *prices*. Redeemable bank credit issued to commercial borrowers cannot inflate prices when the market is in a quiescent state. The banking system is passive, except in a speculative state of the market (*cf.* Fullarton and Tooke). Irredeemable paper issued by or for governments is sharply distinguished from commercial credit; it can produce inflation. These points are illustrated by numerous historical incidents.

In the last third of the book an interpretation of English banking history a century ago is offered for its bearing on current policy. The crisis of 1825 compelled the rejection both of the official bank view that notes issued against solid commercial paper could not depreciate and of Ricardo's theory that redeemable paper currency is automatically kept at its purchasing power parity by free gold movements. But subsequently Ricardo's theory of *metallic* currency became the official view in the restricted note issue established by Peel's Act. Previous crises should have shown the folly of such a restriction. It caused havoc in the crisis of 1847. But with occasional suspensions it continued for 84 years because the Ricardian theory failed to treat as currency, and so similarly to restrict, deposits. (Ricardo's neglect of the fact that gold, unlike other commodities, has a fixed price, and of the importance of interest rates for gold movements is noted.) The moral drawn from these facts and from an examination of the Currency-School-Banking-School controversy is the impracticability of an automatic, mechanical credit policy.

The author's use of statistics may fairly be described as "second hand," but this does not detract from his meritorious attempt to make his theory fit the

facts. In his discussion of recent credit policy it is perhaps to be regretted that he does not deal with the dilemma of a central bank, desiring to check an orgy of security speculation and to avoid both the draining of gold from other central banks and the checking of "the smooth unobstructed movement of goods" which are not overstocked. The passivity of a banking system in a quiescent market is perhaps overstated. And yet in spite of the author's imperfect mastery of English idiom, he offers a compact, coherent formulation of credit theory that no one interested in the subject can afford to neglect.

MORRIS A. COPELAND

Cornell University

Commission Internationale pour la Revision Décenniale des Nomenclatures Internationales des Maladies (causes de décès, causes d'incapacité de travail) devant servir á l'establissement des statistiques nosologiques (classification Bertillon). Quatrieme Session, 16–19 Octobre, 1929. Paris. Imprimerie Nationale. 1930.

The text of the proceedings of the fourth meeting convened during October, 1929, by the Government of France to consider needed revisions in the International List of Causes of Death became available in the United States in June, 1930. The International List in its present general form was originally submitted to the meeting of the International Statistical Institute in Chicago in 1893 by the late Dr. Jacques Bertillon. It was approved for adoption in the United States by the meeting of the American Public Health Association in 1898 and was adopted in other countries at various times between 1894 and 1899. As shown by the proceedings of the International Statistical Institute, the problem of titular uniformity in the international statistics of causes of death was of prime concern to the Institute since its foundation in Brussels in 1853. In that year the grandfather of Dr. Bertillon, Dr. Achille Guillard, proposed to the Institute the establishment of an international nosonomical list. The first attempt at an international nomenclature of diseases was made under Institute auspices shortly after 1853 by Drs. William Farr and Marc d'Espine. (*See* 16th Annual Report of the Registrar General of England and Wales, 1853, Appendix, p. 73.) The general history of the List may be determined from the introductory text of the American edition of the Manual as published by the Division of Vital Statistics of the Bureau of the Census. There should be prepared for our American students of vital statistics an accurate account of the history and development of these projects for uniform nomenclature and classification of diseases and injuries, beginning perhaps with the Cullenian system. (*See* Eldad Lewis, *A Methodological System of Nosology*, Stockbridge, Massachusetts, 1807.)

The work preparatory to the October, 1929, meeting of the Commission was described by Dr. Michel Huber in the *Bulletin of the International Statistical Institute*, Tome XXI, partie 2, and in Tome XXIII, liv. 2, p. 3 (Cairo meeting). In the spring of 1929 a mixed commission composed of delegates of the International Statistical Institute and the Public Health Section of the League of Nations met in Geneva and drew up a tentative revision of the titles of the List. This report, together with the results of other preliminary inquiries in individual

countries, was of use and value to the Commission when it met in Paris during October, 1929. An extensive memorandum and numerous individual suggestions were submitted by the American delegation.

The result of the Commission's deliberations appears on pages 89 *et seq.* of the present report. Offices of vital statistics are offered: I. The detailed Nomenclature of Causes of Death (200 titles); II. The Intermediate Nomenclature (85 titles); III. The Abridged Nomenclature (43 titles); IV. Nomenclature of Causes of Stillbirths (13 titles); V. Supplementary Morbidity Nomenclature (finer subdivision of 12 titles of the International List for purposes of sickness statistics); VI. Nine recommendations for the promotion of uniformity in statistical classification and presentation (standard age groups, certification of deaths, treatment of jointly returned causes of death, use of the nomenclature, statistical disposition of deaths following surgical operation, deaths following industrial injuries, and the importance of morbidity statistics).

The "convention" signifying general acceptance of the International List was signed by the representatives of 33 countries (with Australia, British India and Canada included under British Empire). Signatories of international organizations represented the International Labor Bureau (Nixon), the International Statistical Institute (Delatour), the League of Red Cross Societies (Sand), the International Public Hygiene Office (Abt) and the Hygiene Office, League of Nations (Stouman, Park).

An aid to the use of the List is shown in pages 111 to 163, giving some of the major returns of causes of death to be included under each of the titles. A little dictionary of terms used in reporting diseases and conditions (pp. 165–222) is of further assistance in promoting uniformity. The American edition of the List is at present under way in the office of Dr. Timothy F. Murphy, Chief Statistician for Vital Statistics, Bureau of the Census.

In certain respects the country-wide adoption of the 1929 revision of the List will result in the appearance of a number of new titles and subtitles in the published mortality tables. Tuberculosis of the skin and subcutaneous cellular tissue, of the lymphatic system and of the genito-urinary system will be shown separately. Important changes were made in the segregation of cancers, in the classification of heart and circulatory diseases, and in the classification of diseases related to pregnancy, childbirth and the puerperal state generally. The Commission definitely rejected the American proposal to make obligatory the subclassification of accidental deaths on the basis of whether death occurred in the home, in industry or in public places.

The American delegation attending the Conference consisted of Dr. Timothy F. Murphy (Chairman); Mr. George H. Van Buren, Drs. William H. Guilfoy, W. J. V. Deacon, Emlyn Jones, Rupert Blue, Robert Pierret and Miss Jessamine S. Whitney.

E. W. Kopf

The Small Loan Situation in New Jersey in 1929, by Willford Isbell King. Trenton: The New Jersey Industrial Lenders' Association. 1929. 101 pp.

This study should be read in conjunction with the report made by Pace, Gore and MacLaren entitled *Financial Aspects of the Small Loan Situation in New Jersey*, which, I feel, should have been published under the same cover. Pace, Gore and MacLaren treat the subject of expense and profit of the various operating licensees; while Dr. King purposely devotes himself to a discussion of the borrowers in their various aspects.

The study was undertaken at the request of the New Jersey Industrial Lenders' Association. Since 1914 Small Loan Companies in New Jersey had been operating under a law that permitted them to charge a rate of 3 per cent per month on unpaid balances on loans not in excess of $300 to any one borrower. This rate had been thought excessive, and a new bill reducing the maximum interest rate to 1½ per cent per month was before the State Legislature at the time the report was written. The bill was passed with the result that only about 20 per cent of the previous licensees renewed their licenses; and it is understood that only about two-thirds of those renewing intend to operate.

Dr. King's exhaustive study includes such tabulations as the average size of loan compared with the estimated monthly family income, the purposes for which money was borrowed, average size of loans, and a further breaking up of borrowers into occupational classes.

The report, complete with tables and charts, is exceedingly well written, and shows more thought for the human side of the question than is usually found in statistical literature.

JOHN G. ROLPH

The Amalgamated Bank of New York

Statistical Methods for Research Workers, by R. A. Fisher. Third edition. Edinburgh: Oliver and Boyd. 1930. xiii, 283 pp. and folded tables.

The first two editions of this valuable book were reviewed in this JOURNAL for September, 1927, and September, 1928. In addition to an extension of the simplified method of fitting parabolic trends, two sections have now been added which elucidate respectively the maximum-likelihood estimation of parameters from fragmentary data, and Dr. Fisher's philosophy of "amount of information." This last idea, which he uses to replace the unsatisfactory methods of inverse probability in measuring the accumulation of knowledge, is novel and striking.

The amount of information which a sample contains relative to the value of a parameter p is the negative of the second derivative with respect to p of the logarithm of the likelihood. The likelihood may be defined as the probability that this sample should occur. Hence doubling the size of the sample doubles the amount of information. The efficiency of any statistic is the ratio of the amount of information which it uses effectively to the total amount in the sample. The standard error is the square root of the reciprocal of the amount of information used. Most efficient of all statistics is that value of p for which the likelihood is a maximum. If we use this function of the observations we shall conserve all

our information, some of which leaks away when we rely on less efficient statistics, though these are often easier to compute.

A genetical experiment of Tedin's is shown to have yielded 3725 units of information. Of these units, 31.77 per plant were derived from one part of the experiment, 35.58 from another, and only .06 from a third. Since the land and labor depend on the number of plants, which may be varied at will, a cost accounting system for such work is thus made possible. Efficiency engineering in the production of knowledge on a large scale seems to be at hand, with units of information economized like calories and kilowatt hours.

HAROLD HOTELLING

Stanford University

Joint Occupation Study. The Actuarial Society of America and The Association of Life Insurance Medical Directors. 1928.

This statistical investigation was undertaken as the result of a suggestion made in a paper presented in 1927 to the Actuarial Society of America which pointed out that the latest collective statistics on occupational ratings for life insurance policies was from 19 to 42 years old. Much of the information given in the Medico-Actuarial Mortality Investigation was not believed suitable for present-day application, since many industries had changed materially since the World War. The investigation was made possible by the uniform occupation code proposed in 1926 which has been generally adopted. A copy of the code appears in the appendix and contains provision for more than 1,000 different occupations.

The data used in the survey were furnished by twelve of the larger life insurance companies of this country, the companies comprising three-fifths of the total insurance in force in the United States. There are two groups of classes by years of issue, 1915 to 1926 and 1920 to 1926, carried to the anniversaries of the policies in 1927. The investigation is based upon observations by policies, both men and women being included. All policies which would have been classed sub-standard because of build, race, residence, or medical impairment are excluded. No Industrial, Group, or Intermediate ($500 policies) are included.

Two basic tables, one for each group of years, were prepared. In each case the function graduated was the ratio of the rate of mortality to that of the American Men Select Table. Since the durations of the tables, twelve years and seven years, are short, it was not deemed advisable to make a calculation of the ultimate rate of mortality.

The experience under each occupation or group of occupations is listed and compared to the expected number of deaths according to the basic table. For the convenience of those who may wish to avail themselves of the figures as a guide in gauging the significance of the data, there is given, following the mortality ratios of the report, the value of the approximation to the probable deviation $\pm \frac{2}{3}/\sqrt{\theta}$, where θ is the expected number of death claims in the particular class.

Of particular interest in the report is a "Table of Deaths by Cause as Percentage of Total Deaths" showing the distribution of the deaths entering into each of

the basic tables. A table showing the double accidental death experience by occupation is also valuable. The principal classes in which there is a considerable difference between the average of company ratings and the mortality experienced are listed in a very interesting way.

The authors of the report have used considerable care in describing in detail the results of the investigation in each class and their suggestions are valuable in interpreting these results. The details of the report are such that it would be impossible to include them in this brief review and can only be appreciated by a careful study of the comments of the Committee in the report itself.

WALTER O. MENGE

University of Michigan

An Appraisal of American Business Forecasts, by Garfield V. Cox. University of Chicago Studies in Business Administration, Volume 1, No. 2. 1929. 88 pp.

In a sense this book is a sequel to *Forecasting Business Conditions* (Macmillan, 1927), which Cox wrote in collaboration with C. O. Hardy. The readers of this earlier book may remember that an appendix contained key phrases typical of the monthly forecasts of four forecasting services for the period November, 1918, to August, 1925. Extending the time to produce a full ten-year period, and expanding the forecasting services to six, Cox has appraised the key phrases by two tests. The six services now included are Babson's, Brookmire, Harvard, Moody's, National City Bank, and Standard.

Test No. 1 is a quantitative measure of how well the forecasts of each month during the ten-year period forecast the subsequent event. Test No. 2 is a qualitative and quantitative test of how well the seven major turning points (changes of 12 or more points in the Annalist Index of Business Activity) were forecasted.

In Test No. 1 grades are assigned for definiteness and correctness. The product gives the score for adequacy. The two sets of grades are built on a scale such that there are 17 possible scores for adequacy, ranging from -1 to $+1$ by intervals of $\frac{1}{4}$. The grade assigned for *definiteness* depends upon how clearly were stated the time to which the forecast was supposed to apply and the character of the forecasted event. The grade assigned for *correctness* is decided by appealing to the probable results of the given policy rather than to other actual or ideal forecasts. The criterion used is one of four different indexes of general business activity. Concerning *correctness* Cox states, "Where the most general prediction concerned only a limited field of activity, the forecast was nevertheless checked solely against subsequent developments in that special field" (pp. 19–20, *cf.* pp. 11, 14).

By the application of Test No. 1 the services scored on the average $+0.66$ for *definiteness*, $+0.53$ for *correctness*, and $+0.36$ for *adequacy*. The services scoring the highest were incorrect about 1 month out of 10; those scoring lowest were incorrect about 1 month out of 4.

In Test No. 2 grades are assigned for anticipating (1) the direction of the major turn, (2) the amplitude of the major turn, and (3) the time at which the

major turn began. The score for adequacy is obtained by multiplying the first of these grades by the sum of the second and third. The scales are so built up that the possible scores for adequacy range from -1 to $+1$ by intervals of $1/6$. In this test, one proceeds from the major turn back to the forecast. Scoring is begun at the earliest month in which any of the services actually succeeded in anticipating the turn and continues slightly beyond the date of the turn.

Cox does not emphasize the numerical size of the scores in Test No. 2. The scores are classified as follows: helpful, slightly helpful, neutral, slightly misleading, and misleading. Out of 40 cases (a case being one service forecasting one major turn) 13 are helpful and 18 slightly helpful. In making this tabulation, Cox gives some consideration to the time at which individual monthly scores were made and to the consistency of sign of these individual scores. No statement is made as to what the critical points in the averages are, but apparently $+.25$ tends to mark a break between the two highest classes and $-.25$ to mark a break between the two lowest classes.

Scoring by 11 students in both Test No. 1 and Test No. 2 gave results similar to those obtained by Cox himself.

Under the caption, "The Relation of Forecasting Methods to Results," Cox gives a succinct statement of the basic methods pursued by each one of the services. This should be of value not only as a ready reference, but also as a spur to forecasters to study the methods pursued by others, especially since it is presented in the same book with invidious comparisons between services. One item may be especially noted. Cox found in reconstructing parts of the old Babson Barometer by Babson methods that there was ridiculous dispersion between the series averaged in terms of the Babson chart scale (pp. 59–60).

In appendices, Cox presents tables giving all the separate grades assigned.

Just what does the size of the scores indicate? Zero means a service is neither right nor wrong. What does a score of say $+0.30$ mean? There seems to be no clear basis for interpreting this as a 30 per cent or yet as a 65 per cent adequate forecast. Is the value of the scores, besides indicating whether the service is right or wrong, to give a basis of comparison between one forecast and another? If so, what should we say of a forecast which rates $+.45$ as compared to one which rates $+.30$? Merely that the former is better? Is a score of $+0.60$ as much better than $+0.45$ as $+0.45$ is better than $+0.30$? How large must the difference in score be, before it is significant? (See pages 24 and 39 where it is indicated that a difference of 6 points on a base of $+0.26$ is not very significant in Test No. 1, but that a difference of 11 points on a similar base is; and that a difference of 8 points on a base of $+.18$ is not very significant in Test No. 2.)

Can we compare directly the size of the score in Test No. 1 with the size of the score in Test No. 2? The two grades are combined in Test No. 1 by multiplication while the two major grades in Test No. 2 are combined by addition. If the grading is done with equal rigidity, this tends to make the scores larger in Test No. 2. Is there any reason except convenience for making the assumptions implicit in these methods of combining grades?

The criterion used in Test No. 2 is the Annalist Index of Business Activity. It should be noted that the services studied were not unanimously agreed that

this is a better measure of general business than a price or value series. The Harvard Business Curve is an average of a wholesale price index and bank debits (p. 47); Brookmire used Bradstreet's index of commodity prices as representative of general business for most of the ten-year period covered (pp. 52–53); Standard used bank debits as representative of general business in the years 1921–1923 (p. 54); during most of the period Babson Service included price series in its measure of general business (pp. 59–60), and Moody's Trade Barometer used regularly throughout the period includes price and value series (p. 65). None of these series were first deflated.

Cox recognizes that the particular rating assigned any given service is scarcely an adequate basis for forecasting the relative efficiency of the given forecasting service in the future (p. 69). But of considerable importance is his finding that "the group as a whole should average more right than wrong."

ELMER CLARK BRATT

Lehigh University